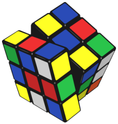
Splittings of (block) Hessenberg M-matrices

Luca Gemignani, Federico Poloni
Department of Computer Science, University of Pisa



ALAMA/2GGALN Conference
Alcalá de Henares, June ~~2020~~2022

M-Matrix splittings

Let A be an M-matrix: for simplicity in this talk $A = I - P$, with $P \geq 0$ and $\rho(P) < 1$.

A regular splitting is a decomposition $P = P_1 + P_2$, with $P_1, P_2 \geq 0$.

Classical topic associated to iterative methods for linear systems $Ax = b$:

$$Ax = b \implies (I - P_1)x = P_2 x + b \implies (I - P_1)x_{k+1} = P_2 x_k + b.$$

Equivalently: Richardson iteration preconditioned with $I - P_1$:

$$x_{k+1} = x_k + (I - P_1)^{-1}(b - Ax_k).$$

$P_1 = \mathrm{diag}(P)$	Jacobi
$P_1 = \mathrm{tril}(P)$	Gauss-Seidel
$P_1 = \mathrm{triu}(P)$	anti-Gauss-Seidel

Other splittings

Other splittings are useful, especially if they are permuted triangular $\to$ systems $(I - P_1)x_{k+1} = \dots$ can be solved by substitution.

Staircase splitting

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & & & & \\ & * & * & * & & \\ & & & * & & \\ & & & * & * & * \\ & & & & & * \end{bmatrix}, \quad P_2 = \begin{bmatrix} & & * & * & * & * \\ * & & * & * & * & * \\ * & & & & * & * \\ * & * & * & & * & * \\ * & * & * & & & \\ * & * & * & * & * & \end{bmatrix}.$$

One can compute first all even-numbered entries of x_{k+1} in parallel, then all odd-numbered ones.

This makes sense in a parallel setting, especially blockwise ($* =$ blocks).

Comparison theorems

The asymptotic convergence speed of the iterative method $(I - P_1)x_{k+1} = P_2 x_k + b$ is given by the spectral radius $\rho((I - P_1)^{-1} P_2) =: \mathrm{sr}(P_1)$, natural comparison measure.

Various classical results exist.

Theorem [Varga '61, Woźnicki '01]

Given two splittings P_1, P_2 and $\hat{P}_1, \hat{P}_2$,

- $\hat{P}_1 \geq P_1 \implies \mathrm{sr}(\hat{P}_1) \leq \mathrm{sr}(P_1)$
- $A^{-1}\hat{P}_2 A^{-1} \leq A^{-1} P_2 A^{-1} \implies \mathrm{sr}(\hat{P}_1) \leq \mathrm{sr}(P_1)$
- $\hat{P}_2 A^{-1} \leq (A^{-1} P_2)^T \implies \mathrm{sr}(\hat{P}_1) \leq \mathrm{sr}(P_1)$

Idea: The larger P_1 is, the closer $I - P_1$ is to $A = I - P_1 - P_2 \to$ more effective preconditioner.

Comparing splittings

These theorems allow one to compare effectiveness.

Example (Stein–Rosenberg theorem)

$\mathrm{diag}(P) \leq \mathrm{tril}(P) \implies \mathrm{sr}(\text{Gauss–Seidel}) \leq \mathrm{sr}(\text{Jacobi})$

Many splittings cannot be compared:

Example

No inequalities hold in general between $\mathrm{triu}(P)$ and $\mathrm{tril}(P)$ $\implies$ Gauss–Seidel and anti-Gauss–Seidel are incomparable.

Comparing the incomparable

Our main result looks like it should be in a 1980s book (but we looked around and didn't find it).

Theorem [Gemignani, P. '21]

For a lower Hessenberg M-matrix,

$$\mathrm{sr}(\text{anti-Gauss–Seidel}) \leq \mathrm{sr}(\text{staircase}) \leq \mathrm{sr}(\text{Gauss–Seidel}).$$

More generally, anti-Gauss–Seidel beats any splitting with a permuted triangular $P_1 = \Pi$ ◣ Π^*.

Analogous results hold for upper Hessenberg matrices (replacing GS with AGS), and blockwise.

A counterintuitive result

$$A = \begin{bmatrix} * & {\color{red}*} & & & & \\ {\color{blue}*} & * & {\color{red}*} & & & \\ {\color{blue}*} & {\color{blue}*} & * & {\color{red}*} & & \\ {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & * & {\color{red}*} & \\ {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & * & {\color{red}*} \\ {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & {\color{blue}*} & * \end{bmatrix}$$

Black + red is a better preconditioner than black + blue.

Even with fewer and possibly smaller entries, e.g., in

$$A = \begin{bmatrix} 1 & {\color{red}-\varepsilon} & 0 \\ {\color{blue}-100} & 1 & {\color{red}-\varepsilon} \\ {\color{blue}-100} & {\color{blue}-100} & 1 \end{bmatrix}.$$

Why does this case matter?

$$A = \begin{bmatrix} * & * & & & & \\ * & * & * & & & \\ * & * & * & * & & \\ * & * & * & * & * & \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{bmatrix}$$

Several queuing theory problems result in block Hessenberg matrices: Markov chains with skip-free 'levels', e.g., number of customers in a queue.

Typically the blocks themselves are sparse, making them prime candidates for iterative methods. [Latouche–Ramaswami book '99; Dudin, Dudin et al. '20]; also [Meini '97, Bini–Latouche–Meini book '05] for the infinite-dimensional case.

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & & & & & \\ {\color{red}*} & * & & & & \\ {\color{red}*} & * & * & & & \\ {\color{red}*} & * & * & * & & \\ {\color{red}*} & * & * & * & * & \\ {\color{red}*} & * & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix}P_{11} & P_{12}\\ 0 & P_{22}\end{bmatrix}, \begin{bmatrix}0 & 0\\ P_{21} & 0\end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix}P_{11} & 0\\ P_{21} & P_{22}\end{bmatrix}, \begin{bmatrix}0 & P_{12}\\ 0 & 0\end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & & & & \\ & * & * & & & \\ & * & * & * & & \\ & * & * & * & * & \\ & * & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\mathrm{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \mathrm{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & & & & \\ {\color{red}*} & * & * & & & \\ {\color{red}*} & * & * & * & & \\ {\color{red}*} & * & * & * & * & \\ {\color{red}*} & * & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & & & \\ & & * & * & & \\ & & * & * & * & \\ & & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & & & \\ * & * & * & * & & \\ * & * & * & * & * & \\ * & * & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\mathrm{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \mathrm{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & {\color{red}*} & & \\ & & & * & & \\ & & & * & * & \\ & & & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & & \\ * & * & * & * & * & \\ * & * & * & * & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\operatorname{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \operatorname{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & {\color{red}*} & \\ & & & & * & \\ & & & & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\mathrm{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \mathrm{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & * & \\ & & & & * & \\ {\color{red}*} & {\color{red}*} & {\color{red}*} & {\color{red}*} & * & * \end{bmatrix}$$

Proof

The swap lemma

$$\mathrm{sr}\left(\begin{bmatrix} P_{11} & P_{12} \\ 0 & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ P_{21} & 0 \end{bmatrix}\right) = \mathrm{sr}\left(\begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix}, \begin{bmatrix} 0 & P_{12} \\ 0 & 0 \end{bmatrix}\right)$$

Proof: same characteristic polynomial $\det((I - P_1)x - P_2)$.

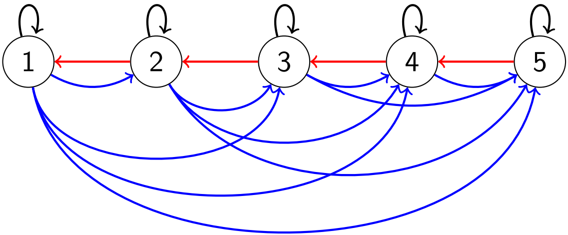
We get from GS (or other splittings) to AGS with a sequence of moves that are either (1) swap lemma applications or (2) increasing entries of P_1.

$$P_1 = \begin{bmatrix} * & * & & & & \\ & * & * & & & \\ & & * & * & & \\ & & & * & * & \\ & & & & * & * \\ & & & & & * \end{bmatrix}$$

A combinatorial proof

Sketch of a combinatorial proof that $\operatorname{sr}(AGS) \leq \operatorname{sr}(GS)$:

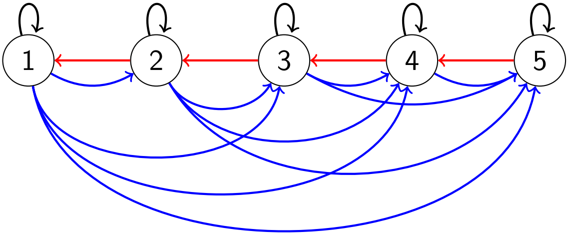
- Consider P as the adjacency matrix of a directed weighted graph. The splitting $P = P_1 + P_2$ partitions edges into $E = E_1 \sqcup E_2$.
- Main idea: if the adjacency matrix of a graph is lower Hessenberg, red edges (from i to $j < i$) appear in a walk at least as often as blue ones, since we can only decrease by one at a time: $i \to i-1 \to i-2 \to \dots$.



$$\begin{bmatrix} * & * & & & \\ * & * & * & & \\ * & * & * & * & \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}$$

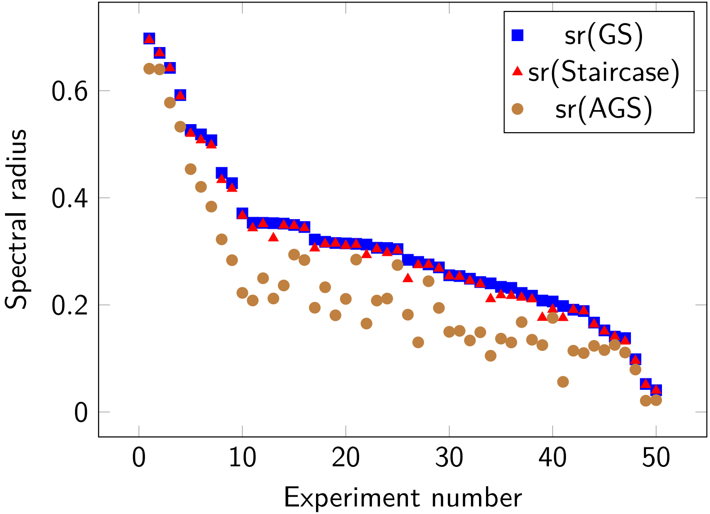
A combinatorial proof – cont.

- Matrix multiplication $\leftrightarrow$ counting walks: $P_1 P_2$ counts walks with one step in E_1 and then one in E_2.
- $R = (I - P_1)^{-1} P_2 = (I + P_1 + P_1^2 + \dots) P_2$ counts walks with an arbitrary number of E_1-steps and end with an E_2-step.
- In GS, R^k counts walks with k red edges; in AGS, it counts walks with k blue edges.
- This argument gives an inequality $R^k_{AGS} \lesssim R^k_{GS}$.



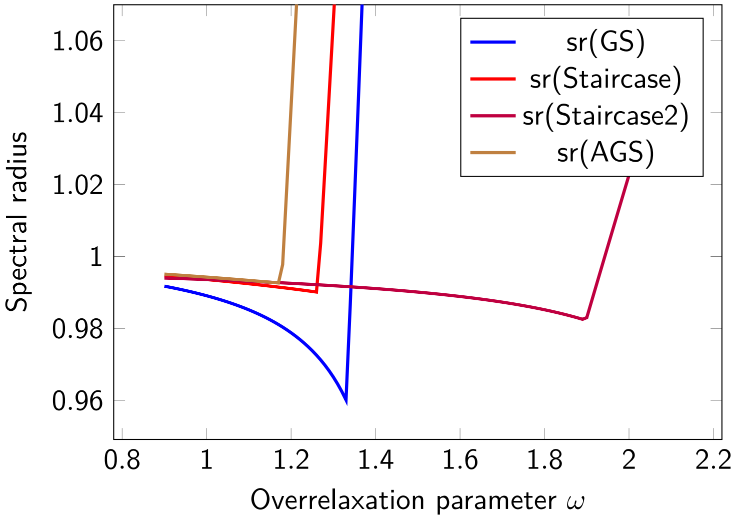
$$\begin{bmatrix} * & * & & & \\ * & * & * & & \\ * & * & * & * & \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix}$$

Experiments



Iteration radii for 50 random 5×5 lower Hessenberg M-matrices, sorted by decreasing value of $\rho(GS)$.

Experiments



Iteration radii with over-relaxation, matrix Q from [Dudin, Dudin et al '20] (upper block Hessenberg with 20 blocks of size 48 each).

Conclusions

- Little counterintuitive result that somehow eluded the 1980s.
- Potential for queuing theory applications.
- Some insight to drive preconditioner/splitting choices.

Thanks for your attention!

Gemignani, P. *Comparison theorems for splittings of M-matrices in (block) Hessenberg form*. BIT 2021.