Eigenvalue estimates for saddle point matrices arising in weak constraint four-dimensional variational data assimilation

Ieva Daužickaitė

University of Reading

Amos Lawless, Jennifer Scott, Peter Jan van Leeuwen

24 November, 2020
Sparse Days

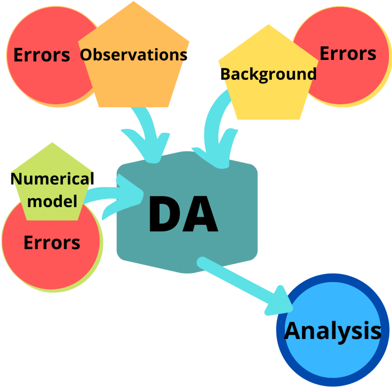
Objective

Explore the sensitivity of the spectra of system matrices in weak constraint four-dimensional variational data assimilation.

We will look at:

- Theory on how the extreme eigenvalues change when new observations are introduced.
- Eigenvalue bounds.
- Numerical example to illustrate the theory.

State formulation of weak constraint 4D-Var



$$J(x_0,\ldots,x_N) = \frac{1}{2}||x_0 - x^b||_{B^{-1}} + \frac{1}{2}\sum_{i=0}^{N}||y_i - \mathcal{H}_i(x_i)||_{R_i^{-1}} + \frac{1}{2}\sum_{i=0}^{N-1}||x_{i+1} - \mathcal{M}_i(x_i)||_{Q_{i+1}^{-1}},$$

where $x_{i+1} = \mathcal{M}_i(x_i) + \eta_{i+1}$, $\eta_i \sim N(0, Q_i)$, $x_i \in \mathbb{R}^n$.

Incremental weak constraint 4D-Var

Search for the state $\mathbf{x} \in \mathbb{R}^{n(N+1)}$ incrementally (equivalent to a Gauss-Newton method):

Update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}$, where $\delta\mathbf{x}$ minimises quadratic cost function

$$J^{\delta}(\delta\mathbf{x}^{(k)}) = \frac{1}{2}||\mathbf{L}^{(k)}\delta\mathbf{x}^{(k)} - \mathbf{b}^{(k)}||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\mathbf{H}^{(k)}\delta\mathbf{x}^{(k)} - \mathbf{d}^{(k)}||^2_{\mathbf{R}^{-1}},$$

- $\mathbf{L}^{(k)}$ and $\mathbf{H}^{(k)}$ includes the linearised model and observation operator, respectively;
- $\mathbf{b}^{(k)}$ includes the model errors;
- $\mathbf{d}^{(k)}$ includes the discrepancy between the state and the observations.

Incremental weak constraint 4D-Var

Search for the state $\mathbf{x} \in \mathbb{R}^{n(N+1)}$ incrementally (equivalent to a Gauss-Newton method):

Update $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \delta\mathbf{x}^{(k)}$, where $\delta\mathbf{x}$ minimises quadratic cost function

$$J^{\delta}(\delta\mathbf{x}^{(k)}) = \frac{1}{2}||\mathbf{L}^{(k)}\delta\mathbf{x}^{(k)} - \mathbf{b}^{(k)}||^2_{\mathbf{D}^{-1}} + \frac{1}{2}||\mathbf{H}^{(k)}\delta\mathbf{x}^{(k)} - \mathbf{d}^{(k)}||^2_{\mathbf{R}^{-1}},$$

- $\mathbf{L}^{(k)}$ and $\mathbf{H}^{(k)}$ includes the linearised model and observation operator, respectively;
- $\mathbf{b}^{(k)}$ includes the model errors;
- $\mathbf{d}^{(k)}$ includes the discrepancy between the state and the observations.

$\delta\mathbf{x}$ can be found by solving large-sparse linear systems of equations.

We will consider **saddle point** and **SPD** systems.

3×3 block saddle point formulation

Fisher and Gürol (2017) propose obtaining $\delta \mathbf{x}$ by solving

$$\mathcal{A}_3 \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \delta \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{0} \end{pmatrix},$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are Lagrange multipliers, and

$$\mathcal{A}_3 = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^T & \mathbf{H}^T & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(2(N+1)n+p)\times(2(N+1)n+p)},$$

n - number of model variables, N - number of time steps, p - total number of observations, $p << n(N+1)$.

3×3 block saddle point formulation

Fisher and Gürol (2017) propose obtaining $\delta \mathbf{x}$ by solving

$$\mathcal{A}_3 \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{\mu} \\ \delta \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{0} \end{pmatrix},$$

$$\mathbf{L} = \begin{pmatrix} I & & & \\ -M_0 & I & & \\ & \ddots & \ddots & \\ & & -M_{N-1} & I \end{pmatrix}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are Lagrange multipliers, and

$$\mathcal{A}_3 = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^T & \mathbf{H}^T & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{(2(N+1)n+p)\times(2(N+1)n+p)},$$

n - number of model variables, N - number of time steps, p - total number of observations, $p << n(N+1)$.

2×2 block saddle point formulation

We also propose obtaining $\delta\mathbf{x}$ by solving

$$\mathcal{A}_2 \begin{pmatrix} \boldsymbol{\lambda} \\ \delta\mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ -\mathbf{H}^T\mathbf{R}^{-1}\mathbf{d} \end{pmatrix},$$

where

$$\mathcal{A}_2 = \begin{pmatrix} \mathbf{D} & \mathbf{L} \\ \mathbf{L}^T & -\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \end{pmatrix} \in \mathbb{R}^{2(N+1)n \times 2(N+1)n}.$$

1×1 block formulation

$\delta\mathbf{x}$ can also be obtained by solving the standard formulation

$$\mathcal{A}_1 \delta\mathbf{x} = \mathbf{L}^T \mathbf{D}^{-1}\mathbf{b} + \mathbf{H}^T \mathbf{R}^{-1}\mathbf{d},$$

where

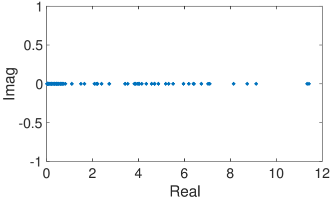
$$\mathcal{A}_1 = (\mathbf{L}^T \mathbf{D}^{-1}\mathbf{L} + \mathbf{H}^T \mathbf{R}^{-1}\mathbf{H}) \in \mathbb{R}^{(N+1)n \times (N+1)n}$$

is SPD.

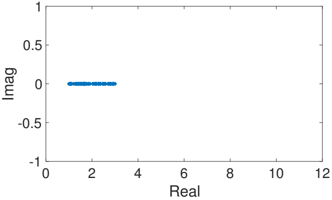
Krylov subspace solvers and eigenvalues

- Krylov solvers require preconditioning for satisfactory performance.
- Previous research on preconditioning 3×3 block system gave disappointing results (Fisher and Gurol (2017), Freitag and Green (2018), Gratton et al (2018)).
- The rate of convergence of Krylov subspace iterative solvers for symmetric systems depends on the spectrum of the coefficient matrix.

"Troublesome" spectrum



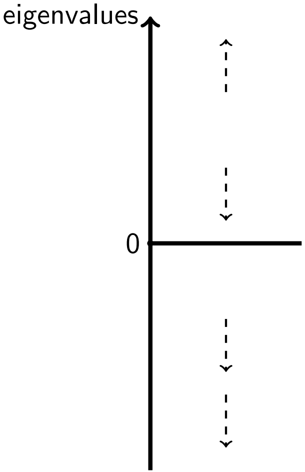
"Nice" spectrum



- To explore how the observations influence the convergence, we look into how the spectra of the saddle point and SPD matrices change when new observations are introduced.

Change of extreme eigenvalues of $\mathcal{A}_3$

$$\mathcal{A}_3 = \begin{pmatrix} \mathbf{D} & \mathbf{0} & \mathbf{L} \\ \mathbf{0} & \mathbf{R} & \mathbf{H} \\ \mathbf{L}^T & \mathbf{H}^T & \mathbf{0} \end{pmatrix}$$

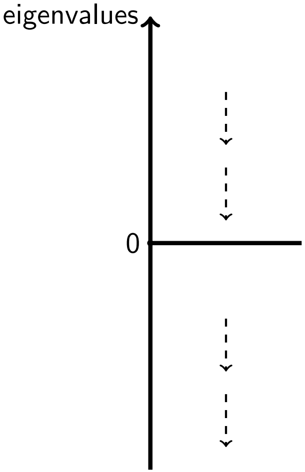


Theorem

The smallest and largest negative eigenvalues of $\mathcal{A}_3$ either move away from the origin or are unchanged when new observations are introduced. The same holds for the largest positive eigenvalue, while the smallest positive eigenvalue approaches the origin or is unchanged.

Change of extreme eigenvalues of $\mathcal{A}_2$

$$\mathcal{A}_2 = \begin{pmatrix} \mathbf{D} & \mathbf{L} \\ \mathbf{L}^T & -\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \end{pmatrix}$$

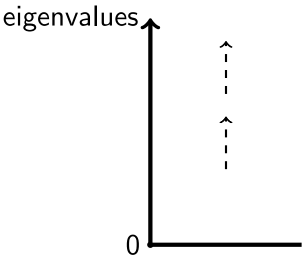


Theorem

If the observation errors are uncorrelated, i.e. $\mathbf{R}$ *is* **diagonal**, *then the smallest and largest negative eigenvalues of* $\mathcal{A}_2$ *either move away from the origin or are unchanged when new observations are added. Contrarily, the smallest and largest positive eigenvalues of* $\mathcal{A}_2$ *approach the origin or are unchanged.*

Change of extreme eigenvalues of $\mathcal{A}_1$

$$\mathcal{A}_1 = \mathbf{L}^T \mathbf{D}^{-1} \mathbf{L} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$$



Theorem

If the observation errors are uncorrelated, i.e. $\mathbf{R}$ *is* **diagonal**, *then the eigenvalues of* $\mathcal{A}_1$ *move away from the origin or are unchanged when new observations are added.*

Eigenvalue bounds with fixed number of observations

- Bounds for eigenvalues of $\mathcal{A}_3$ and $\mathcal{A}_1$ depend on:
  - extreme eigenvalues of $\mathbf{D}$;
  - extreme eigenvalues of $\mathbf{R}$;
  - largest and smallest nonzero singular values of $(\mathbf{L}^T \quad \mathbf{H}^T)$.
- Bounds for eigenvalues of $\mathcal{A}_2$ depend on:
  - extreme eigenvalues of $\mathbf{D}$;
  - extreme eigenvalues of $\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$;
  - largest eigenvalue of $\mathbf{R}$;
  - extreme singular values of $\mathbf{L}$;
  - the smallest nonzero singular value of $(\mathbf{L}^T \quad \mathbf{H}^T)$.

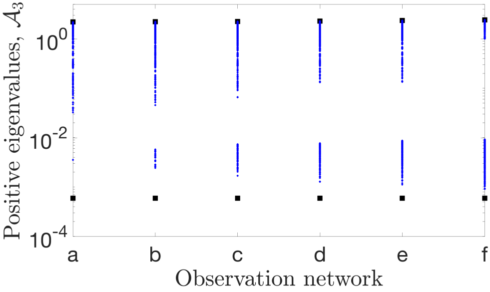
Numerical example. Assimilation system

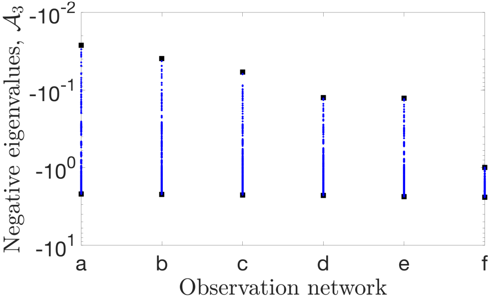
- Lorenz 96 model where the evolution of variable X^k, $k \in \{1, 2, \ldots, n\}$, is governed by the set of n coupled ODEs:

$$\frac{dX^k}{dt} = -X^{k-2}X^{k-1} + X^{k-1}X^{k+1} - X^k + F.$$

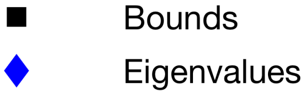
- $n = 40$, $F = 8$, $N = 15$, $\Delta t = 0.025$, $\Delta x = 0.025$.
- $B_i = Q_i = 0.05^2 C_q$, where C_q is SOAR with correlation length scale 1.5×10^{-2}, $R_i = 0.1^2 I$.
- Direct observations.
- Integrated with 4th order Runge-Kutta scheme.
- Performed with Matlab. Part of the code is written by A. El-Said.

Spectrum of $\mathcal{A}_3$ w.r.t. the number of observations





O.n.	Eigenvalues
a)	$[-2.192, -2.99 \times 10^{-2}]$ $[3.56 \times 10^{-3}, 2.195]$
f)	$[-2.408, -9.96 \times 10^{-1}]$ $[9.14 \times 10^{-4}, 2.413]$



Observation networks:

a) $p = 1$ at the final time,

b) $p = 20$, observing every 8*th* model variable at every 4*th* time step,

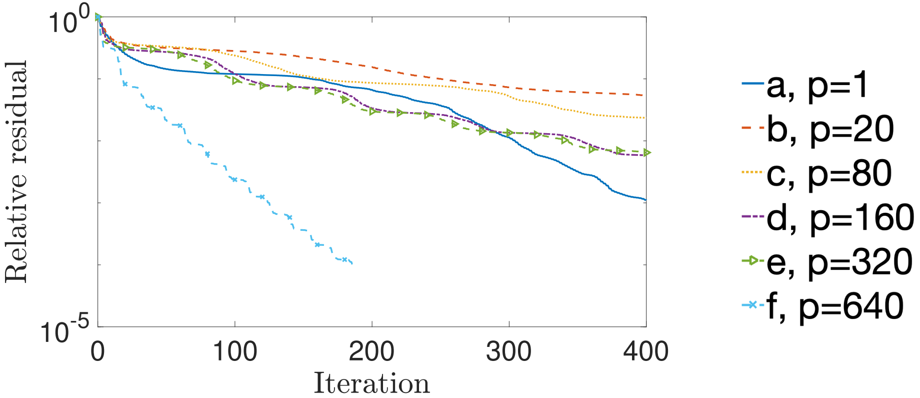
c) $p = 80$, observing every 4*th* model variable at every 2*nd* time step,

d) $p = 160$, observing every 2*nd* model variable at every 2*nd* time step,

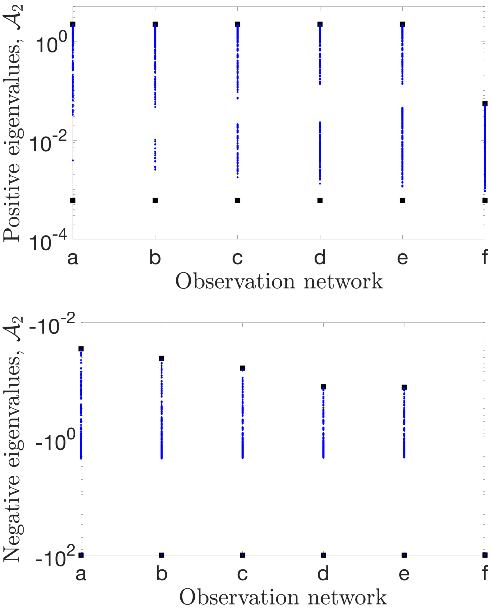
e) $p = 320$, observing every 2*nd* model variable at every time step,

f) $p = 640$, fully observed system.

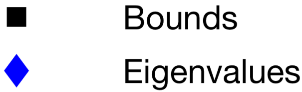
MINRES convergence. 3×3 block system



Spectrum of $\mathcal{A}_2$ w.r.t. the number of observations



O.n.	Eigenvalues
a)	$[-1.0001 \times 10^2, -2.99 \times 10^{-2}]$ $[3.91 \times 10^{-3}, 2.195]$
f)	$[-1.0005 \times 10^2, -1.00 \times 10^2]$ $[9.35 \times 10^{-4}, 5.15 \times 10^{-2}]$



Observation networks:

a) $p = 1$ at the final time,

b) $p = 20$, observing every 8th model variable at every 4th time step,

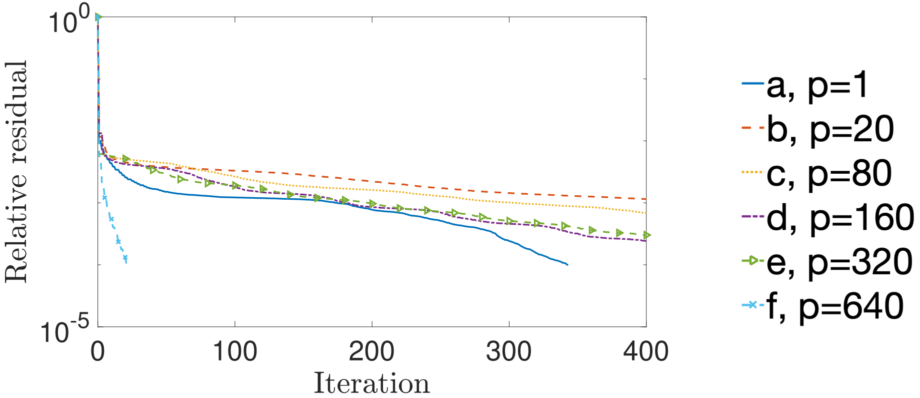
c) $p = 80$, observing every 4th model variable at every 2nd time step,

d) $p = 160$, observing every 2nd model variable at every 2nd time step,

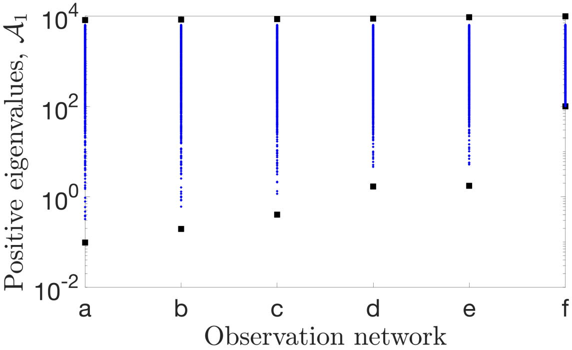
e) $p = 320$, observing every 2nd model variable at every time step,

f) $p = 640$, fully observed system.

MINRES convergence. 2×2 block system



Spectrum of $\mathcal{A}_1$ w.r.t. the number of observations



O.n.	Eigenvalues
a)	$[3.23 \times 10^{-1}, 6.30 \times 10^3]$
f)	$[1.00 \times 10^2, 6.40 \times 10^3]$

■ Bounds

◆ Eigenvalues

Observation networks:

a) $p = 1$ at the final time,

b) $p = 20$, observing every $8th$ model variable at every $4th$ time step,

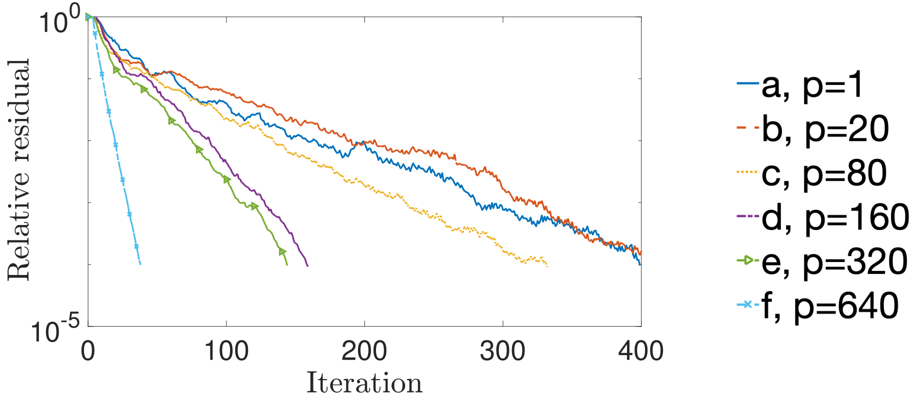
c) $p = 80$, observing every $4th$ model variable at every $2nd$ time step,

d) $p = 160$, observing every $2nd$ model variable at every $2nd$ time step,

e) $p = 320$, observing every $2nd$ model variable at every time step,

f) $p = 640$, fully observed system.

CG convergence. 1×1 block system



Conclusions

- We have formulated theorems on how the extreme eigenvalues of $\mathcal{A}_3$, $\mathcal{A}_2$ and $\mathcal{A}_1$ change w.r.t. new observations.
  - Extreme negative eigenvalues of $\mathcal{A}_3$ and $\mathcal{A}_2$, largest positive eigenvalue of $\mathcal{A}_3$, and extreme eigenvalues of $\mathcal{A}_1$ move away from the origin. Smallest positive eigenvalue of $\mathcal{A}_3$ and extreme positive eigenvalues of $\mathcal{A}_2$ move towards the origin.
- We have determined bounds for the eigenvalues of $\mathcal{A}_3$, $\mathcal{A}_2$ and $\mathcal{A}_1$.
- Numerical experiments illustrate theoretical sensitivity analysis, and show that the spectral intervals are tight for $\mathcal{A}_3$ and $\mathcal{A}_2$, but pessimistic for $\mathcal{A}_1$.
- The small positive eigenvalues of $\mathcal{A}_2$ and $\mathcal{A}_3$ can cause convergence issues when new observations are added.
- Including the information on observations coming from the observation error covariance matrix $\mathbf{R}$ and the linearised observation operator $\mathbf{H}$ could be beneficial for preconditioning.

References

I. Daužickaitė, A. S. Lawless, J. A. Scott, and P. J. van Leeuwen
Spectral estimates for saddle point matrices arising in weak constraint four-dimensional variational data assimilation, Numerical Linear Algebra with Applications, 27(5):e2313, 2020.

M. Fisher and S. Gürol
Parallelisation in the time dimension of four-dimensional variational data assimilation, Quarterly Journal of the Royal Meteorological Society, 143(703):1136 - 1147, 2017.

M. A. Freitag and D. L. H. Green
A low-rank approach to the solution of weak constraint variational data assimilation problems, Journal of Computational Physics, 357:263 - 281, 2018.

S. Gratton, S. Gürol, E. Simon, and Ph. L. Toint
Guaranteeing the convergence of the saddle formulation for weakly-constrained 4D-VAR data assimilation, Quarterly Journal of the Royal Meteorological Society, 144(717):2592 - 2602, 2018.