Adjoints and low-rank covariance representation

Michael K. Tippett

International Research Institute for climate prediction, LDEO, Palisades, NY

Stephen E. Cohn, Ricardo Todling

Data Assimilation Office, NASA/GSFC, Greenbelt, MD, USA.

Dan Marchesin

Instituto de Matemática Pura e Aplicada, Rio de Janeiro, RJ, Brazil

Motivation

Want estimates *and* characterizations of their error.

- Ensemble prediction – analysis and analysis errors.
- Data assimilation – first-guess and first-guess errors.

Question

When are low-dimension error covariance models appropriate?

Approach

Look at error covariance matrix equation and linear operators on covariance matrices. Operator norms and adjoints.

Conclusion

Low-rank representations related to ill-conditioned error covariance equation.

Familar ideas

Discrete-time setting

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k$$

Inner product on vectors,

$$(\mathbf{x}, \mathbf{x}) = \mathbf{x}^T \mathbf{x}$$

leads to operator adjoints,

$$(\mathbf{x}, \mathbf{A}_k \mathbf{x}) = (\mathbf{A}_k^T \mathbf{x}, \mathbf{x})$$

and operator norms,

$$\|\mathbf{A}_k\|^2 = \max_{\mathbf{x}} \frac{\|\mathbf{A}_k \mathbf{x}\|^2}{\|\mathbf{x}\|^2} = \max_{\mathbf{x}} \frac{|\mathbf{x}^T \mathbf{A}_k^T \mathbf{A}_k \mathbf{x}|}{|\mathbf{x}^T \mathbf{x}|}$$

Apply same formalism to linear operators on error covariance matrices.

Familar ideas?

Inner product on matrices

$$(\mathbf{X}, \mathbf{Y}) = \operatorname{tr} \mathbf{X}^* \mathbf{Y}$$

leads to operator adjoints

$$(\mathbf{X}, \mathcal{L}\mathbf{Y}) = (\mathcal{L}^* \mathbf{X}, \mathbf{Y})$$

Matrix norms lead to operator norms

$$\|\mathcal{L}\|_p = \max_{\mathbf{X}} \frac{\|\mathcal{L}\mathbf{X}\|_p}{\|\mathbf{X}\|_p}$$

Matrix p-norms

$$\|\mathbf{X}\|_p^p = \sum_{i=1}^{n} \sigma_i^p(\mathbf{X})$$

For $\mathbf{P} \geq 0\,, \quad \|\mathbf{P}\|_1 = \operatorname{tr} \mathbf{P}\,, \quad \|\mathbf{P}\|_\infty = \lambda_1(\mathbf{P})$.

Simple $\mathcal{L}$

Discrete-time setting

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k + \mathbf{b}_{k+1}$$

$$\left\langle \mathbf{x}_k \mathbf{x}_k^T \right\rangle = \mathbf{P}_k \,, \quad \left\langle \mathbf{b}_k \mathbf{b}_j^T \right\rangle = \delta_{kj} \mathbf{Q}$$

Time-independent steady-state:

$$\mathcal{L}\mathbf{P} \equiv \mathbf{P} - \mathbf{A}\mathbf{P}\mathbf{A}^T = \mathbf{Q}$$

What properties of $\mathcal{L}$ help answer:

- When do a few modes explain much of the variance of $\mathbf{P}$?
- How can imperfect models of $\mathbf{Q}$ be used?

What is special about $\mathcal{L}$?

$\mathcal{L}\mathbf{P} \equiv \mathbf{P} - \mathbf{A}\mathbf{P}\mathbf{A}^T$

- $\mathbf{P} = \mathcal{L}^{-1}\mathbf{Q}$ exists for stable $\mathbf{A}$.
- If $\mathbf{Q}$ is a covariance matrix, so is $\mathbf{P} = \mathcal{L}^{-1}\mathbf{Q}$.

$\boxed{\mathcal{L}^{-1} \text{ is a positive map.}}$

Positive maps are nice.

(Farrell, Kleeman).

$\mathcal{L}\mathbf{P} = \mathbf{Q}$.

Maximum response is to uncorrelated forcing.

$$\|\mathcal{L}^{-1}\|_\infty = \max_{\mathbf{Q}} \frac{\lambda_1(\mathcal{L}^{-1}\mathbf{Q})}{\lambda_1(\mathbf{Q})} = \max_{\mathbf{Q}} \frac{\lambda_1(\mathbf{P})}{\lambda_1(\mathbf{Q})} = \lambda_1(\mathcal{L}^{-1}\mathbf{I})$$

Maximum variance.

$$\|\mathcal{L}^{-1}\|_1 = \max_{\mathbf{Q}} \frac{\operatorname{tr}\mathcal{L}^{-1}\mathbf{Q}}{\operatorname{tr}\mathbf{Q}} = \|\mathcal{L}^{-*}\|_\infty = \lambda_1(\mathcal{L}^{-*}\mathbf{I}) = \|\mathcal{L}^{-1}\mathbf{w}\mathbf{w}^T\|_1$$

$\mathbf{w}$ is first EOF of $\mathcal{L}^{-*}\mathbf{I}$.

EOFs of $\mathcal{L}^{-*}\mathbf{I}$ order state-space according to variance produced.

Positive maps are nice (2).

Conditioning of $\mathcal{L}$ is related to low-rank representations of $\mathbf{P}$.

$\kappa(\mathcal{L}) = \|\mathcal{L}\|_\infty \|\mathcal{L}^{-1}\|_\infty$ is the condition number of $\mathcal{L}$.

$$\frac{\lambda_1(\mathbf{P})}{\lambda_n(\mathbf{P})} \le \kappa(\mathcal{L}) \frac{\lambda_1(\mathbf{Q})}{\lambda_n(\mathbf{Q})}$$

For $\mathbf{Q} = \mathbf{I}$, $\mathbf{P}$ has good low-rank approximation only when $\mathcal{L}$ is ill-conditioned.

$\mathcal{L}\mathbf{P} \equiv \mathbf{P} - \mathbf{A}\mathbf{P}\mathbf{A}^T$ is ill-conditioned if $\mathbf{A}$ is *almost* unstable.

- $\mathbf{A}$ has almost unstable eigenmodes.
- $\mathbf{A}$ has large singular values.
- $\mathbf{A}$ has sensitive eigenvalues.

Positive maps are nice (3).

Operator norm of $\mathcal{L}$ related to to low-rank representations of $\mathbf{P}$.

$$\frac{\lambda_1(\mathbf{P})/\operatorname{tr}\mathbf{P}}{\lambda_1(\mathbf{Q})/\operatorname{tr}\mathbf{Q}} \leq \|\mathcal{L}^{-1}\|_\infty \|\mathcal{L}\|_1$$

Most of the variance of $\mathbf{P}$ is in the first EOF if

- the same is true of $\mathbf{Q}$, or
- $\|\mathcal{L}^{-1}\|_\infty \|\mathcal{L}\|_1$ is large.

Positive maps are nice (4).

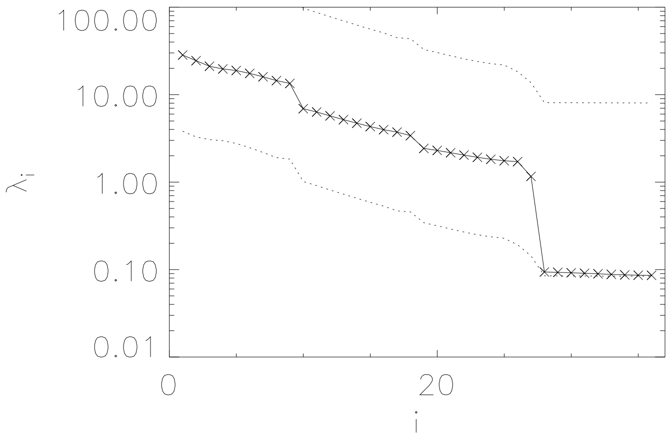
Preserve ordering.

Upper and lower estimates for **Q**,

$$\mathbf{Q}^- \le \mathbf{Q} \le \mathbf{Q}^+$$

imply upper and lower estimates for **P**.

$$\mathcal{L}^{-1}\mathbf{Q}^- \le \mathbf{P} \le \mathcal{L}^{-1}\mathbf{Q}^+$$



$$\mathrm{tr}\,\mathcal{L}^{-1}\mathbf{Q}^- \le \mathrm{tr}\,\mathbf{P} \le \mathrm{tr}\,\mathcal{L}^{-1}\mathbf{Q}^+\,, \qquad \lambda_i(\mathcal{L}^{-1}\mathbf{Q}^-) \le \lambda_i(\mathbf{P}) \le \lambda_i(\mathcal{L}^{-1}\mathbf{Q}^+)$$

Generalizes to time-dependent problems.

Application: Data assimilation

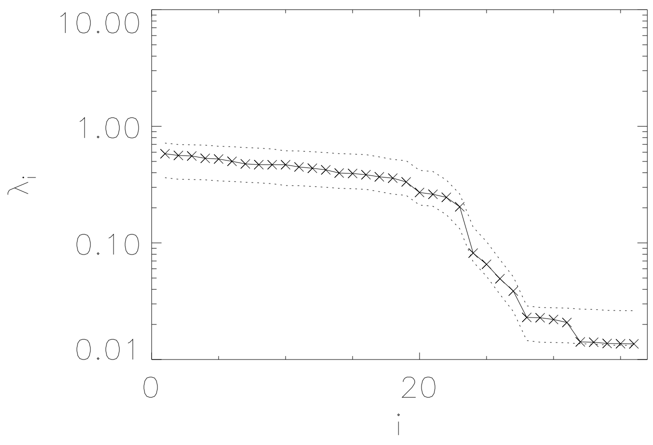
P – is the steady-state analysis error covariance.

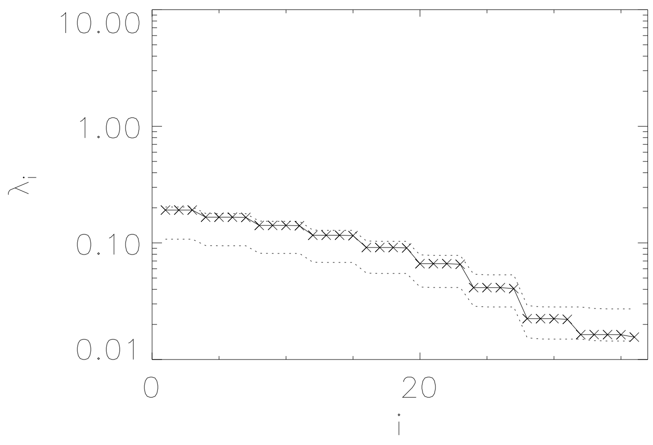
A – combines time-independent dynamics *and* analysis.

Q – contains model and observational error covariances.

Not observing modes that *grew*, increases the singular values of **A**.

Forces upper and lower bounds for analysis error to increase.





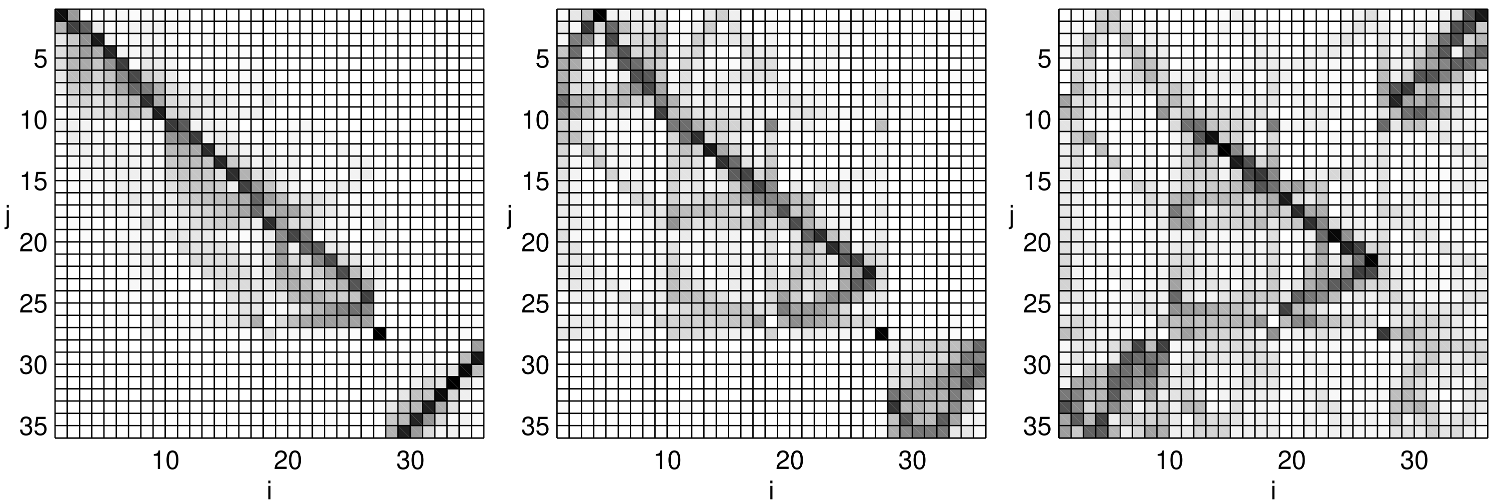
Did not observe modes that grew .

Did observe modes that grew.

Estimating EOFs of $\mathbf{P}$

Projecting eigenvectors of analysis error covariance $\mathbf{P}$ onto eigenvectors of:

Estimate ($\mathbf{Q} = \mathbf{I}$) $\qquad$ $\mathbf{A}\mathbf{A}^T$ $\qquad$ $\mathbf{A}^T\mathbf{A}$



Estimates of EOFs of $\mathbf{P}$ are good when
bounds are tight *and* $\mathbf{P}$ has low-rank approximation.

Conclusions

- Matrix operator and their adjoints are useful tools–theoretical understanding and estimates.
- Can be calculated using SVD and stochastic optimal machinery.