Stability of the Spectral EnKF under nested covariance estimators

Marie Turčičová*[1], Jan Mandel[2], and Kryštof Eben[3]

[1] *Charles University in Prague, Faculty of Mathematics and Physics, Sokolovská 83, Prague 8, 186 75, Czech Republic*
[2] *University of Colorado Denver, Denver, CO 80217-3364, USA*
[3] *Czech Academy of Sciences, Institute of Computer Science, Pod Vodárenskou věží 271/2, 182 07 Praha 8, Czech Republic*

In the case of traditional Ensemble Kalman Filter (EnKF), it is known that the filter error does not grow faster than exponentially for a fixed ensemble size [5]. The question posted in this contribution is whether the upper bound for the filter error can be improved by using an improved covariance estimator that comes from the right parameter subspace and has smaller asymptotic variance. Its effect on Spectral EnKF is explored by a simulation.

Keywords: nested covariance models, maximum likelihood, error of EnKF

1 Introduction

Estimating of large covariance matrices from small samples is an important problem in many fields, including spatial statistics, genomics, and ensemble filtering. One of the prominent applications is data assimilation, where a prior estimate of a random vector (usually representing a system state) is adjusted in order to be more consistent with current observations. The revised estimate is then plugged into a time-evolution model as an initial condition for the future time prediction. This approach, known as filtering, is used in many fields including meteorological predictions. A characteristic feature of this application is a large dimension of the system state (millions or larger), which results in high computational cost. One algorithm that deals with this problem is the Ensemble Kalman filter (EnKF), which approximates the mean and the covariance of the state vector from an ensemble. However, due to the high computational cost, this ensemble is always very small compared to the state dimension, and the approximation is very poor. In this contribution, we study improved estimation of the covariance matrix from a small ensemble, and its behaviour in high-dimensional EnKF.

*turcic@karlin.mff.cuni.cz

In particular, we consider a very special type of sparse approximation of a covariance matrix in spectral space, based on nested maximum likelihood models for diagonal matrices. The improved covariance estimator seems to have a positive effect in data assimilation, which is illustrated by a simulation.

2 Hierarchical maximum likelihood estimators

Suppose $\mathbb{X}_N = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N]$ is a random sample from a distribution on $\mathbb{R}^n$ with density $f(\boldsymbol{x}, \boldsymbol{\theta})$ with unknown parameter vector $\boldsymbol{\theta}$ in a parameter space $\Theta \subset \mathbb{R}^p$. The maximum likelihood estimator (MLE) $\hat{\boldsymbol{\theta}}_N$ of the true parameter $\boldsymbol{\theta}^0$ is defined by maximizing the log-likelihood $\ell(\boldsymbol{\theta}|\mathbb{X}_N) = \sum_{i=1}^{N} \log f(\boldsymbol{X}_i, \boldsymbol{\theta})$.

Further assume a hierarchical structure of the parameter space,

$$\boldsymbol{\theta}^0 \in \Psi \subset \Phi \subset \Theta,$$

where $\Psi \subset \mathbb{R}^m$, $\Phi \subset \mathbb{R}^k$, $m \le k \le p$. That is, $\boldsymbol{\theta}$ can be parametrized by a smaller number of parameters. We assume that the map $\boldsymbol{\varphi} \mapsto \boldsymbol{\theta}(\boldsymbol{\varphi})$ is one-to-one from Φ to Θ and continuously differentiable. Further assume that the associated Jacobi matrix $\nabla_{\boldsymbol{\varphi}} \boldsymbol{\theta}(\boldsymbol{\varphi}) = \left\{ \frac{\partial \theta_i}{\partial \varphi_j} \right\}$ has full rank for all $\boldsymbol{\varphi} \in \Phi$. We make analogous assumptions about the map $\boldsymbol{\psi} \mapsto \boldsymbol{\theta}(\boldsymbol{\psi})$ as well. Moreover, assume that $\boldsymbol{\theta}^0 = \boldsymbol{\theta}(\boldsymbol{\varphi}^0) = \boldsymbol{\theta}(\boldsymbol{\psi}^0)$ is an interior point of Ψ.

We will also adopt the usual assumptions in the maximum likelihood theory: (i) the density f determines the parameter $\boldsymbol{\theta}$ uniquely in the sense that $f(\boldsymbol{x}, \boldsymbol{\theta}_1) = f(\boldsymbol{x}, \boldsymbol{\theta}_2)$ a.e. if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, and (ii) $f(\boldsymbol{x}, \boldsymbol{\theta})$ is a sufficiently smooth function of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ (see [6] for details).

Under these assumptions, the error of the estimates is asymptotically normal

$$\sqrt{N}\left(\boldsymbol{\theta}\left(\hat{\boldsymbol{\varphi}}_N\right) - \boldsymbol{\theta}^0\right) \xrightarrow{d} \mathcal{N}_p\left(\mathbf{0}, Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)}\right) \text{ as } N \to \infty, \tag{1}$$

$$\sqrt{N}\left(\boldsymbol{\theta}\left(\hat{\boldsymbol{\psi}}_N\right) - \boldsymbol{\theta}^0\right) \xrightarrow{d} \mathcal{N}_p\left(\mathbf{0}, Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)}\right) \text{ as } N \to \infty. \tag{2}$$

The matrices $Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)}$ and $Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)}$ represent asymptotic variances of the parameters. These matrices are singular, but they can be understood as inverses of Fisher information matrices in a generalized sense. Their exact forms are given in [7].

The next theorem shows that for any two nested subspaces Φ and Ψ of the parameter space containing the true parameter, the asymptotic covariance matrices of the MLE are ordered in the same way. Hence, by confining the parameters to a smaller subspace, we can only improve the estimator.

Theorem 1 ([7]). *Under the assumptions listed previously, the matrix* $Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)} - Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)}$ *is positive semidefinite (denoted as* $Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)} \leq Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)}$*).*

In addition, if $U \sim \mathcal{N}_p\left(\mathbf{0}, Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)}\right)$ *and* $V \sim \mathcal{N}_p\left(\mathbf{0}, Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)}\right)$ *are random vectors with the asymptotic distributions of the estimates* $\boldsymbol{\theta}\left(\hat{\boldsymbol{\varphi}}_N\right)$ *and* $\boldsymbol{\theta}(\hat{\boldsymbol{\psi}}_N)$*, then*

$$\mathrm{E}\,|U|^2 = \frac{1}{N}\,\mathrm{Tr}\,Q_{\boldsymbol{\theta}(\boldsymbol{\varphi}^0)} \leq \frac{1}{N}\,\mathrm{Tr}\,Q_{\boldsymbol{\theta}(\boldsymbol{\psi}^0)} = \mathrm{E}\,|V|^2\,, \tag{3}$$

where $|V| = \left(V^\top V\right)^{1/2}$ *is the standard Euclidean norm in* $\mathbb{R}^p$*.*

2.1 One specific hierarchical model for a covariance matrix

Consider three particular nested models for a diagonal covariance matrix. Such models appear to be useful in meteorological practice but more about our motivation will be said in the next section. The models have the form

- $D^{(n)} = \mathrm{diag}\{d_i, i = 1, \ldots, n\}$
- $D^{(3)} = \mathrm{diag}\{(c_1 - c_2\lambda_i)^{-1}(-\lambda_i)^{-\alpha}, i = 1, \ldots, n\}$
- $D^{(2)} = \mathrm{diag}\{c(-\lambda_i)^{-\alpha}, i = 1, \ldots, n\}$,

with $\{\lambda_i\}_{i=1}^n$ being the eigenvalues of a two-dimensional Laplace operator. The superscripts designate the number of parameters of each model. Under the normality assumption, all these parameters can be estimated from a random sample by the maximum likelihood method. Let $\hat{D}^{(n)}, \hat{D}^{(3)}$ and $\hat{D}^{(2)}$ be the resulting estimates. Notice that $\hat{D}^{(n)}$ is formed simply by the diagonal of sample covariance. The asymptotic hierarchical structure of $\mathrm{cov}(\hat{D}^{(n)})$, $\mathrm{cov}(\hat{D}^{(3)})$ and $\mathrm{cov}(\hat{D}^{(2)})$ is theoretically described in the previous section. The exact form of these estimators and their Fisher information matrices can be found in [7].

However, it is difficult to say something general about the MLEs based on small samples (although they are usually more of interest).

The simulations reported in [7] suggest that the hierarchical structure of the error (3) persists also for small samples. Here we use the hierarchical covariance models in data assimilation.

3 Covariance estimators in data assimilation

Our main objective is to demonstrate the positive effect of the improved covariance estimators $\hat{D}^{(3)}$ and $\hat{D}^{(2)}$ in data assimilation. First, let us briefly recall the Ensemble Kalman Filter (EnKF) [2], in the simple case when the whole state is observed. At the beginning, the distribution of the true state vector $\boldsymbol{X}_t$ is represented by a "forecast ensemble" $\boldsymbol{X}_f^1, \ldots, \boldsymbol{X}_f^N$. The

sample covariance of the forecast ensemble is denoted $\hat{C}_f$. Using the perturbed observations $\boldsymbol{y}^1, \ldots, \boldsymbol{y}^N$ (whose error has covariance R), the forecast ensemble is adjusted and results in an "analysis ensemble" $\boldsymbol{X}_a^1, \ldots, \boldsymbol{X}_a^N$, which is supposed to be "closer" to $\boldsymbol{X}_t$. Its sample covariance is denoted $\hat{C}_a$. The process is governed by the following equations:

$$\boldsymbol{X}_a^j = \boldsymbol{X}_f^j + \hat{C}_f \left(\hat{C}_f + R\right)^{-1} \left(\boldsymbol{y}^j - \boldsymbol{X}_f^j\right) \qquad j = 1, \ldots, N \tag{4}$$

$$\hat{C}_a = \left(I - \hat{C}_f \left(\hat{C}_f + R\right)^{-1}\right) \hat{C}_f. \tag{5}$$

Each member of the analysis ensemble is then pushed forward in time by a function $\eta(\cdot)$, which represents the evolution of the process $\boldsymbol{X}$ in time. This shifted ensemble becomes the forecast, and the whole cycle runs all over again.

It is possible to represent the covariance matrix $\hat{C}_f$ in spectral space [1]. Under the assumption of covariance stationarity, the spectral covariance matrix is diagonal with variances of the coefficients of the expansion of the state in the spectral basis. Filtering methods that take advantage of this result and perform the whole data assimilation process in spectral space, using only the diagonal of spectral sample covariance matrix for $\hat{C}_f$, were studied in [4]. Under the normality assumption, this corresponds to improving the spectral sample covariance by using the maximum likelihood estimator $\hat{D}^{(n)}$ from Subsection 2.1. The question is, whether the filter will perform better when using even more precise estimators like $\hat{D}^{(3)}$ and $\hat{D}^{(2)}$. The improvement can be achieved by searching for the MLE in a correct subspace (or close to it). However, based on climatological data, the power model $\hat{D}^{(2)}$ seems to be reasonable [3].

The critical point of every filtering method is its long-time behaviour and stability, especially for a small ensemble. In the case of traditional EnKF (given by equation (4) and (5)), the filter error does not grow faster than exponentially for a fixed ensemble size [5]. The question is, whether the upper bound for the filter error can be improved by using an improved covariance estimator that comes from the right subspace. This is the subject of our current research. The following simulation suggests that the answer may be positive.

The simulation setting was as follows. First, an initial forecast ensemble of size $N = 5$ and the initial true system state were generated from $\mathcal{N}_n(\mathbf{0}, C)$ with $n = 100$ and $C = FDF^\top$, where F is a Fourier transform and $D = \mathrm{diag}\{c(-\lambda_i)^{-\alpha}, i = 1, \ldots, n\}$ with $c = 50$ and $\alpha = 1.5$. In each cycle, the observations $\boldsymbol{y}^j = \boldsymbol{X}_t + \boldsymbol{\xi}^j$ were generated with $\boldsymbol{\xi}^j \sim \mathcal{N}_n(\mathbf{0}, R)$ and $R = 0.0064 \cdot \mathbb{I}$ and then assimilated with the forecast ensemble. The analysis part was done in the spectral space, following [4], where the theoretical covariance matrix D_f is assumed to be diagonal. After the assimilation part, the analysis ensemble

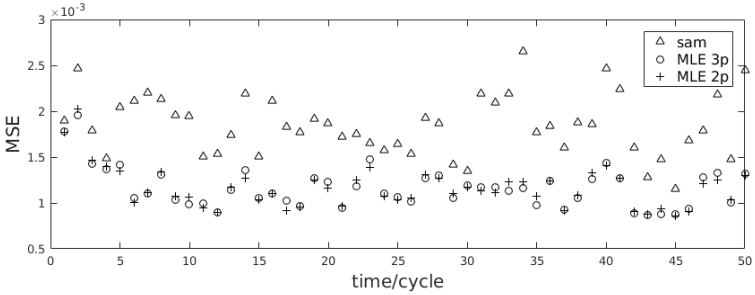

Figure 1: Mean square errors of the analysis ensemble mean.

was propagated in time by the model $\eta(\boldsymbol{X}_a) = A\boldsymbol{X}_a + \boldsymbol{b}$ with $A = 0.9 \cdot \mathbb{I}$ and $\boldsymbol{b} \sim \mathcal{N}_n(\boldsymbol{0}, C)$. The cycle consisting of analysis and propagation step then runs all over again. Three parallel filters were run with distinct estimators of D_f used in the analysis step. The estimators were $\hat{D}_f^{(n)}$ (denoted `sam`), $\hat{D}_f^{(3)}$ (`MLE 3p`) and $\hat{D}_f^{(2)}$ (`MLE 2p`).

In each of 50 cycles, the analysis ensemble was summarized into its mean $\bar{\boldsymbol{X}}_a = \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{X}_a^j$ and the mean square error

$$\frac{1}{n}\sum_{i=1}^{n} \left(\bar{X}_a^j(i) - X_t(i)\right)^2$$

was plotted for every cycle. We denoted by $\bar{X}_a^j(i)$ the entries of $\bar{\boldsymbol{X}}_a^j$. As we can see at Fig. 1, the analysis that uses the more precise covariance estimator is closer to the true state vector (in terms of MSE). However, the performance of the analysis mean is not the only criterion. The stability of the analysis covariance C_a is also important. In Fig. 2, we can see a comparison of spectral representations of four matrices. The true filtering covariance descents from the original covariance C by propagation in time and by assimilation using the expression (5) (where $\hat{C}_f$ is substituted by the matrix C_f resulting from the time-propagation step). The other three matrices are distinct estimates of D_f based on the analysis ensemble after the last cycle. The estimate based on sample covariance is very rough. The MLEs follow the proper trend and provide stable estimates.

This short simulation indicates that the error of the EnKF is smaller when a better covariance estimate is used while the analysis covariance is stable. The theoretical background of this effect is a subject of further research.

Acknowledgements: This work was supported by the grant SVV 2017 No. 260454 and by the U.S. National Science Foundation under grant DMS-1216481.

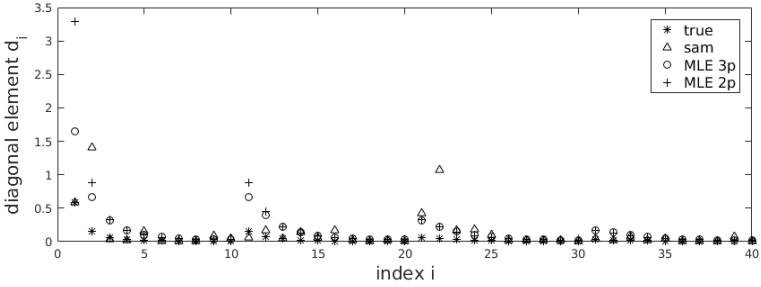

Figure 2: Spectral representations of the true filtering covariance and the analysis covariance matrices (the first 40 elements).

References

[1] P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

[2] G. Evensen. Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99 (C5)(10):143–162, 1994.

[3] G. Gaspari, S. E. Cohn, J. Guo, and S. Pawson. Construction and application of covariance functions with variable length-fields. *Quarterly Journal of the Royal Meteorological Society*, 132(619):1815–1838, 2006.

[4] I. Kasanický, J. Mandel, and M. Vejmelka. Spectral diagonal ensemble Kalman filters. *Nonlinear Processes in Geophysics*, 22(4):485 – 497, 2015.

[5] D. T. B. Kelly, K. J. H. Law, and A. M. Stuart. Well-posedness and accuracy of the ensemble Kalman filter in discrete and continuous time. *Nonlinearity*, 27(10):2579–2603, 2014.

[6] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer-Verlag, New York, second edition, 1998.

[7] M. Turčičová, J. Mandel, and K. Eben. Multilevel maximum likelihood estimation with application to covariance matrices. Technical report. Submitted in January 2017.