Vecchia filters for linear models with Gaussian data

Marcin Jurek

Texas A&M

May 12, 2020

The problem

Let $\mathbf{x}_t$ be the discretized values at time t of the Gaussian process we are interested in.

Using domain knowledge, we also know that

$$\mathbf{x}_t = \mathcal{E}_t(\mathbf{x}_{t-1}) + \mathbf{w}_t, \qquad \text{where } \mathbf{w}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t).$$

where $\mathcal{E}_t$ is some function, possibly non-linear.

The problem

But we don't observe $\mathbf{x}_t$. Instead, we have

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, \qquad \text{where } \mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t).$$

Note that here we assume Gaussian measurement error. This is not necessary in the strict sense, but we will focus on this case for now. The more general case is about to be published.

We want to infer the *filtering distribtution*, i.e.

$$\mathbf{x}_t | \mathbf{y}_{1:t}$$

Traditional methods

When the temporal evolution operator ($\mathcal{E}_t$) is linear (i.e. $\mathcal{E}_t(\mathbf{x}_{t-1}) = \mathbf{E}_t \mathbf{x}_{t-1}$) we normally use the Kalman filter:
Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \boldsymbol{\Sigma}_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{E}_t' + \mathbf{Q}_t$.

Update step:

- Calculate $\mathbf{K}_t := \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t)^{-1}$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_\mathcal{G}} - \mathbf{K}_t \mathbf{H}_t) \boldsymbol{\Sigma}_{t|t-1}$

Traditional methods

Some other methods

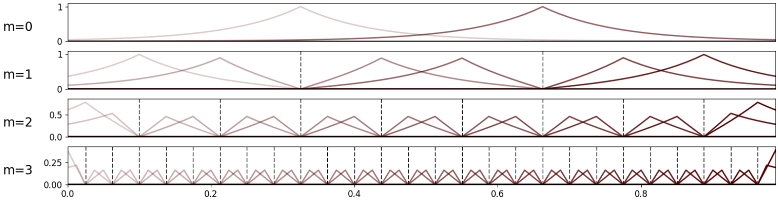
- low-rank filters (LR)
- Ensemble Kalman filter (EnKF)
- filtering in the spectral domain

But: EnKF is using only an ensemble to represent the entire distribution, low-rank filter, is, well... low-rank and (as far as I know) you can do filtering in the spectral domain only in some very special cases.

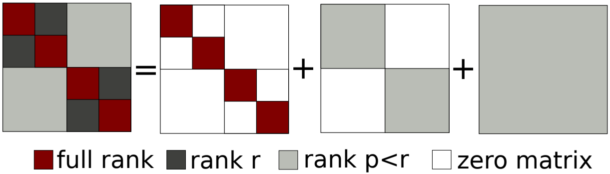
There are so many spatial approximations, what's the problem?

Multi-resolution approximation

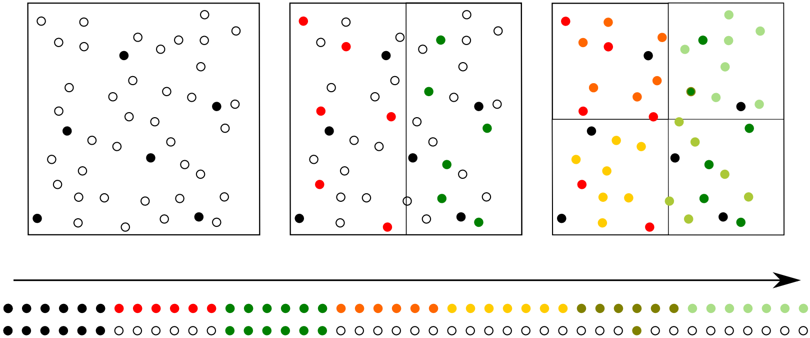
A special basis of functions with decreasing (and "nested") support.



In terms of matrix approximations



MRA as Vecchia

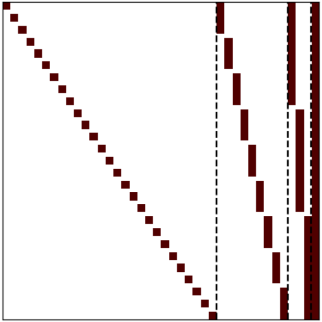


Why is the MRA good for filtering?

Using the basis function approach, MRA enables us to write:

$$\mathbf{x}_t = \mathbf{B}_t \eta_t, \quad \text{where } \eta_t \sim \mathcal{N}(0, \mathbf{I})$$

where $\mathbf{B}_t$ has this special pattern corresponding to the support of each function in the MRA family:



Why is the MRA good for filtering?

What is so special about this structure? Recall the bottleneck in the Kalman filter:

$$(\mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t)^{-1}$$

If we approximate $\mathbf{x}_t \approx \mathbf{B}_t \boldsymbol{\eta}_t$ then, in particular $\mathbf{Var}(\mathbf{x}_t) \approx \mathbf{B}_t \mathbf{B}_t'$ and (dropping time subscripts)

$$\begin{aligned}(\mathbf{H}\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}' + \mathbf{R})^{-1} &\approx (\mathbf{HBBH}' + \mathbf{R})^{-1} = \\ &\mathbf{R}^{-1} + \mathbf{R}^{-1}\mathbf{HB}(\mathbf{I} + \mathbf{H}'\mathbf{B}'\mathbf{RBH})^{-1}\mathbf{B}'\mathbf{H}'\mathbf{R}_t^{-1} \\ &\approx \mathbf{I} + \mathbf{B}(\mathbf{I} + \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' =: \boldsymbol{\Lambda}\end{aligned}$$

Why is the MRA good for filtering

If you do the math and define $\mathbf{L} = \text{chol}(\mathbf{\Lambda})$ then

$$\mathbf{\Sigma}_{t|t} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}' = \mathbf{B}\mathbf{L}\mathbf{L}'\mathbf{B}'$$

Amazingly, $\tilde{\mathbf{B}} := \mathbf{BL}$, has the same sparsity pattern as $\mathbf{B}$.

This means, instead of work with $\mathbf{B}$s instead of $\mathbf{\Sigma}_{t|t}$s.

Let's see how it changes the filtering algo.

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \boldsymbol{\Sigma}_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{E}_t' + \mathbf{Q}_t$.

Update step:

- Calculate $\mathbf{K}_t := \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t)^{-1}$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_{\mathcal{G}}} - \mathbf{K}_t \mathbf{H}_t) \boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}_{0|0}')$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{E}_t' + \mathbf{Q}_t$.

Update step:

- Calculate $\mathbf{K}_t := \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' (\mathbf{H}_t \boldsymbol{\Sigma}_{t|t-1} \mathbf{H}_t' + \mathbf{R}_t)^{-1}$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_{\mathcal{G}}} - \mathbf{K}_t \mathbf{H}_t) \boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t.$

Update step:

- Calculate $\mathbf{K}_t := \boldsymbol{\Sigma}_{t|t-1}\mathbf{H}'_t(\mathbf{H}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}'_t + \mathbf{R}_t)^{-1}$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_{\mathcal{G}}} - \mathbf{K}_t\mathbf{H}_t)\boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t.$

Update step:

- Calculate $\boldsymbol{\Lambda} := (\mathbf{I} + \mathbf{B}'_{t|t-1}\mathbf{H}'_t\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{B}_{t|t-1})^{-1}$.
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_\mathcal{G}} - \mathbf{K}_t\mathbf{H}_t)\boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t.$

Update step:

- Calculate $\boldsymbol{\Lambda} := (\mathbf{I} + \mathbf{B}'_{t|t-1}\mathbf{H}'_t\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{B}_{t|t-1})^{-1}.$
- Calculate $\mathbf{B}_{t|t} := \mathbf{B}_{t|t}\text{chol}(\boldsymbol{\Lambda}_t)$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_\mathcal{G}} - \mathbf{K}_t\mathbf{H}_t)\boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t.$

Update step:

- Calculate $\boldsymbol{\Lambda} := (\mathbf{I} + \mathbf{B}'_{t|t-1}\mathbf{H}'_t \mathbf{R}_t^{-1}\mathbf{H}_t \mathbf{B}_{t|t-1})^{-1}.$
- Calculate $\mathbf{B}_{t|t} := \mathbf{B}_{t|t}\text{chol}(\boldsymbol{\Lambda}_t)$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{B}_{t|t}\mathbf{B}'_{t|t}\mathbf{H}'_t \mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}_t \boldsymbol{\mu}_{t|t-1})$
- Calculate $\boldsymbol{\Sigma}_{t|t} := (\mathbf{I}_{n_\mathcal{G}} - \mathbf{K}_t \mathbf{H}_t)\boldsymbol{\Sigma}_{t|t-1}$

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t.$

Update step:

- Calculate $\boldsymbol{\Lambda} := (\mathbf{I} + \mathbf{B}'_{t|t-1}\mathbf{H}'_t\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{B}_{t|t-1})^{-1}.$
- Calculate $\mathbf{B}_{t|t} := \mathbf{B}_{t|t}\text{chol}(\boldsymbol{\Lambda}_t)$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{B}_{t|t}\mathbf{B}'_{t|t}\mathbf{H}'_t\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1})$
- (x)

Vecchia filter

Assuming that at time $t = 0$ we have

$$\mathbf{x}_0 \sim \mathcal{N}(\boldsymbol{\mu}_{0|0}, \mathbf{B}_{0|0}\mathbf{B}'_{0|0})$$

Forecast step:

- Calculate $\boldsymbol{\mu}_{t|t-1} = \mathbf{E}_t \boldsymbol{\mu}_{t|t}$
- Calculate $\mathbf{B}_{t|t-1} s.t.$
  $\mathbf{B}_{t|t-1}\mathbf{B}'_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1}\mathbf{E}'_t + \mathbf{Q}_t$.

Update step:

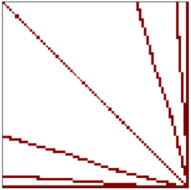
- Calculate $\boldsymbol{\Lambda} := (\mathbf{I} + \mathbf{B}'_{t|t-1}\mathbf{H}'_t\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{B}_{t|t-1})^{-1}$.
- Calculate $\mathbf{B}_{t|t} := \mathbf{B}_{t|t}\text{chol}(\boldsymbol{\Lambda}_t)$
- Calculate $\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{B}_{t|t}\mathbf{B}'_{t|t}\mathbf{H}'_t\mathbf{R}_t^{-1}(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1})$

Why this works

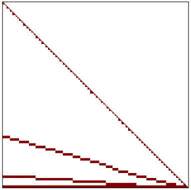
So as you can see it all hinges on the fact that

$$\mathbf{B}_{t|t} = \mathbf{B}_{t|t}\text{chol}(\mathbf{\Lambda}_t)$$

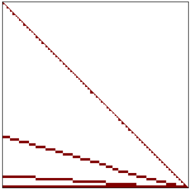
has the same sparsity as $\mathbf{B}_{t|t-1}$ and $\mathbf{B}_{t-1|t-1}$. This is how it looks like:



$\mathbf{I} + \mathbf{B}\mathbf{B}' = \mathbf{\Lambda}^{-1}$



$\mathbf{L} = \text{chol}(\mathbf{\Lambda}^{-1})$



$\mathbf{L}^{-1}$

Why this works

The fact that $\mathbf{L}$ and $\mathbf{L}^{-1}$ have the same sparsity is really special. In fact, it seems like it is the only case of the Vecchia approximation for which it holds.

Most existing spatial approximation methods can be represented as a special case of the general Vecchia approximation. Other than the low-rank filter, we are not aware of any other approximation method which would be applicable to large spatio-temporal data sets.

Extensions

In the pipeline:

- non-linear evolution
- non-Gaussian data using Vecchia-Laplace
- calculating the MRA using incomplete Cholesky decomposition
- extensions to smoothing (go to the end of time and back)

Check out the online extras:

`http://spatial.stat.tamu.edu`