Hierarchical Sparse Cholesky Decomposition with Applications to High-Dimensional Spatio-Temporal Filtering

Marcin Jurek and Matthias Katzfuss

Texas A&M University

August 5, 2020

First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

then

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$$

First step - spatial problem

Naive solution: conditional distributions

Assume

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

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

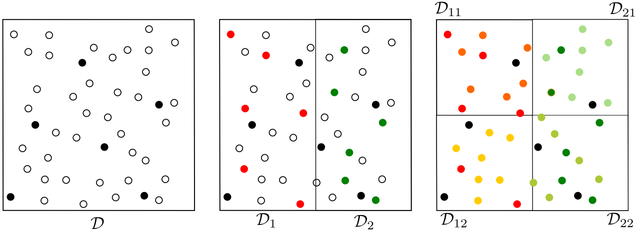
then

$$\mathbf{x}|\mathbf{y} \sim \mathcal{N}(\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\Sigma}})$$

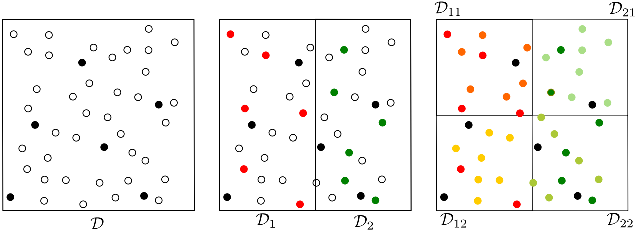
where

$$\begin{aligned}
\boldsymbol{\Lambda}^{-1} &= \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^\top + \mathbf{R} \\
\widetilde{\boldsymbol{\mu}} &= \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{H}^\top\boldsymbol{\Lambda}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}) \\
\widetilde{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{H}^\top\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Sigma}
\end{aligned}$$

The hierarchical Vecchia (HV) approximation

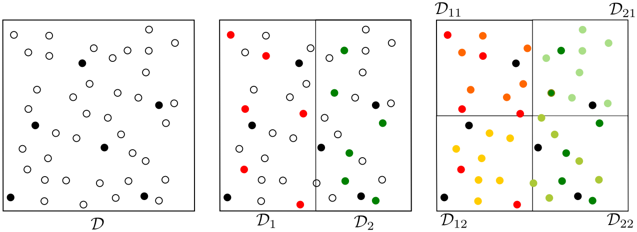


The hierarchical Vecchia (HV) approximation

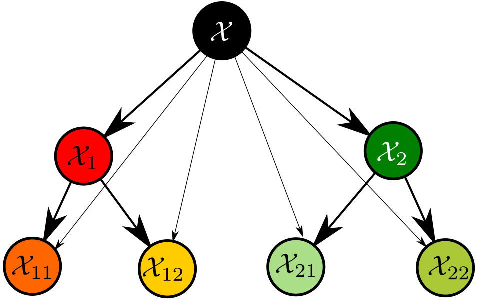


$$\mathbf{x} = \begin{bmatrix} \mathcal{X} \\ \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_{11} \\ \mathcal{X}_{12} \\ \mathcal{X}_{21} \\ \mathcal{X}_{22} \end{bmatrix}$$

The hierarchical Vecchia (HV) approximation



$$\mathbf{x} = \begin{bmatrix} \mathcal{X} \\ \mathcal{X}_1 \\ \mathcal{X}_2 \\ \mathcal{X}_{11} \\ \mathcal{X}_{12} \\ \mathcal{X}_{21} \\ \mathcal{X}_{22} \end{bmatrix}$$



Incomplete Cholesky decomposition

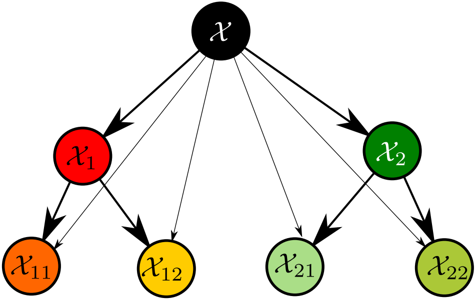
Input: positive-definite matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$, sparsity matrix $\mathbf{S} \in \{0,1\}^{n \times n}$

Result: lower-triangular $n \times n$ matrix $\mathbf{L}$

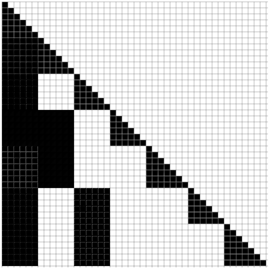
for $i = 1$ to n **do**
  for $j = 1$ to $i - 1$ **do**
    if $\mathbf{S}_{i,j} = 1$ **then**
      $\mathbf{L}_{i,j} = \frac{1}{\mathbf{L}_{j,j}} \left(\mathbf{A}_{i,j} - \sum_{k=1}^{j-1} \mathbf{L}_{i,k} \mathbf{L}_{j,k} \right)$
    end if
  end for
  $\mathbf{L}_{i,i} = \sqrt{\mathbf{A}_{i,i} - \sum_{k=1}^{i-1} \mathbf{L}_{k,k}}$
end for

back to the HV approximation

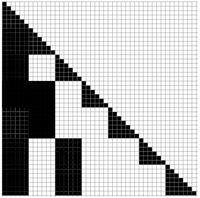
The graph



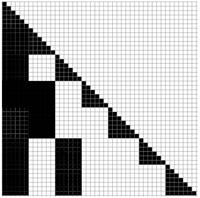
corresponding **S**



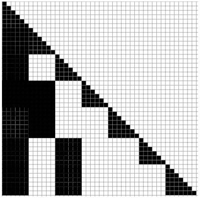
Sparsity patterns (we proved it)



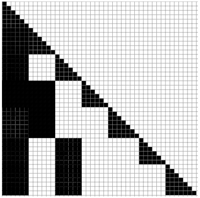
$\mathbf{S}$



$\mathbf{L} := \text{ichol}(\boldsymbol{\Sigma}, \mathbf{S})$



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



$\widetilde{\mathbf{L}} := (\mathbf{P}\,\text{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})\mathbf{P})^{-\top}$

Some matrix algebra

The goal:

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{H}^\top\boldsymbol{\Lambda}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$
$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{H}^\top\boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Sigma}$$

where $\boldsymbol{\Lambda}^{-1} = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^\top + \mathbf{R}$.

Some matrix algebra

The goal:

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{H}^\top \boldsymbol{\Lambda}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$
$$\widetilde{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\mathbf{H}^\top \boldsymbol{\Lambda}\mathbf{H}\boldsymbol{\Sigma}$$

where $\boldsymbol{\Lambda}^{-1} = \mathbf{H}\boldsymbol{\Sigma}\mathbf{H}^\top + \mathbf{R}$.

$$\boldsymbol{\Sigma} \approx \mathbf{L}\mathbf{L}^\top \quad \Longrightarrow \quad \boldsymbol{\Lambda} \approx \mathbf{L}^{-\top}\mathbf{L}^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}$$

$$\left.\begin{array}{c} \mathbf{P} \text{ - reverses order} \\ \widetilde{\mathbf{L}} = (\mathbf{P}\,\mathrm{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})\mathbf{P})^{-\top} \end{array}\right\} \Longrightarrow \quad \widetilde{\boldsymbol{\Sigma}} \approx \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^\top$$

$$\widetilde{\boldsymbol{\mu}} \approx \boldsymbol{\mu} + \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{-\top}\mathbf{H}^\top\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\mu})$$

Posterior inference using HV

Input: $\mathbf{y}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{H}, \mathbf{R}$

Result: $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\mathbf{L}}$

1: $\mathbf{L} = \text{ichol}(\boldsymbol{\Sigma}, \mathbf{S})$
2: $\mathbf{U} = \mathbf{L}^{-\top}$
3: $\boldsymbol{\Lambda} = \mathbf{U}\mathbf{U}^{\top} + \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}$
4: $\widetilde{\mathbf{L}} = \left(\mathbf{P}\left(\text{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})\right)\mathbf{P}\right)^{-\top}$, where $\mathbf{P}$ is the order-reversing permutation matrix
5: $\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{\top}\mathbf{H}^{\top}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}\right)$

Posterior inference using HV

Input: $\mathbf{y}, \mathbf{S}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{H}, \mathbf{R}$

Result: $\widetilde{\boldsymbol{\mu}}$ and $\widetilde{\mathbf{L}}$

1: $\mathbf{L} = \mathrm{ichol}(\boldsymbol{\Sigma}, \mathbf{S})$
2: $\mathbf{U} = \mathbf{L}^{-\top}$
3: $\boldsymbol{\Lambda} = \mathbf{U}\mathbf{U}^{\top} + \mathbf{H}^{\top}\mathbf{R}^{-1}\mathbf{H}$
4: $\widetilde{\mathbf{L}} = \left(\mathbf{P}\left(\mathrm{chol}(\mathbf{P}\boldsymbol{\Lambda}\mathbf{P})\right)\mathbf{P}\right)^{-\top}$, where $\mathbf{P}$ is the order-reversing permutation matrix
5: $\widetilde{\boldsymbol{\mu}} = \boldsymbol{\mu} + \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}^{\top}\mathbf{H}^{\top}\mathbf{R}^{-1}\left(\mathbf{y} - \mathbf{H}\boldsymbol{\mu}\right)$

Can be easily extended to non-Gaussian data using the Vecchia-Laplace algorithm (Zilber & Katzfuss, 2019)

Adding time

Assume that

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

At each time t we want to know the *filtering distribution*:

$$\mathbf{x}_t|\mathbf{y}_{1:t} \sim \mathcal{N}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$$

which is Gaussian, provided that $\mathbf{x}_0|\mathbf{y}_0$ is Gaussian.

Extended Kalman filter

Idea: recursion

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

Assume we know $\mathbf{x}_{t-1} | \mathbf{y}_{1:t-1} \sim \mathcal{N}(\boldsymbol{\mu}_{t-1|t-1}, \boldsymbol{\Sigma}_{t-1|t-1})$

Best guess (forecast):

$$\begin{aligned}
\mathbb{E}(\mathbf{x}_t | \mathbf{y}_{t-1}) &:= \boldsymbol{\mu}_{t|t-1} = \mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1}) \\
\mathbf{E}_t &:= \nabla \mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1}) \\
var(\mathbf{x}_t | \mathbf{y}_{t-1}) &:= \boldsymbol{\Sigma}_{t|t-1} = \mathbf{E}_t \boldsymbol{\Sigma}_{t-1|t-1} \mathbf{E}_t^\top + \mathbf{Q}_t
\end{aligned}$$

Update current data ($\mathbf{y}_t$)
(posterior inference using $\boldsymbol{\mu}_{t|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1}$ as prior moments)

Extended Kalman-Vecchia filter

Input: $\mathbf{S}$, $\boldsymbol{\mu}_{0|0}$, $\boldsymbol{\Sigma}_{0|0}$, $\{(\mathbf{y}_t, \mathcal{E}_t, \mathbf{Q}_t, \mathbf{H}_t, \mathbf{R}_t)\}_{t=1,2,\dots}$,

Result: $\boldsymbol{\mu}_{t|t}$ and $\mathbf{L}_{t|t}$

1: Compute $\mathbf{L}_{0|0} = \big(\mathrm{ichol}(\boldsymbol{\Sigma}_{0|0}, \mathbf{S})\big)^{-\top}$

2: **for** $t = 1, 2, \dots$ **do**

3: Calculate $\mathbf{E}_t = \nabla\mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1})$

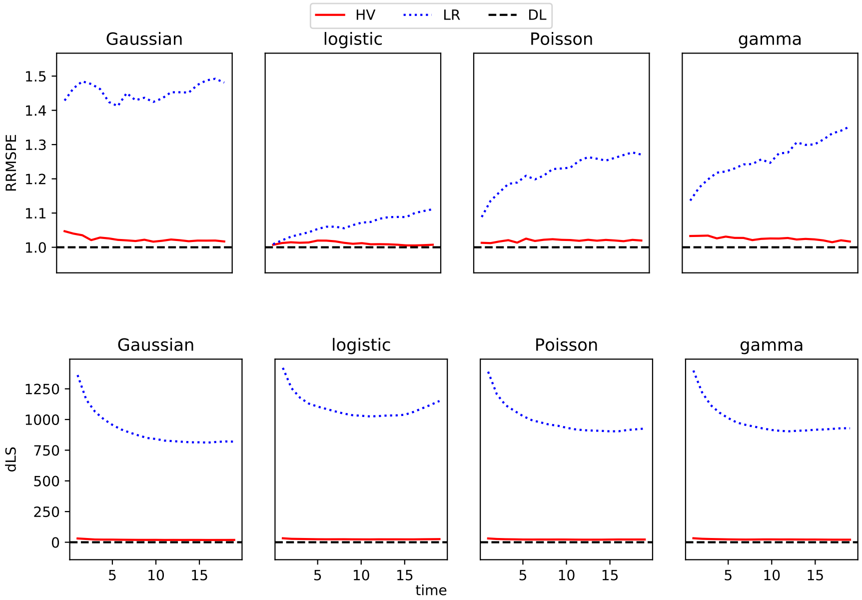
4: Forecast: $$\begin{aligned}\boldsymbol{\mu}_{t|t-1} &= \mathcal{E}_t(\boldsymbol{\mu}_{t-1|t-1}) \\ \mathbf{L}_{t|t-1} &= \mathbf{E}_t\mathbf{L}_{t-1|t-1}\end{aligned}$$

5: Calculate $\boldsymbol{\Sigma}_{t|t-1;i,j} = \mathbf{L}_{t|t-1;i,:}\mathbf{L}^{\top}_{t|t-1;j,:} + \mathbf{Q}_{t;i,j}$ $(i, j : , \mathbf{S}_{i,j} = 1)$:

6: Update: $[\boldsymbol{\mu}_t, \mathbf{L}_{t|t}] = \mathrm{HV}(\mathbf{y}_t, \mathbf{S}, \boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}, \mathbf{H}_t, \mathbf{R}_t)$

7: **return** $\boldsymbol{\mu}_{t|t}, \mathbf{L}_{t|t}$

8: **end for**



HV
LR
DL
Gaussian
logistic
Poisson
gamma
RRMSPE
1.0
1.1
1.2
1.3
1.4
1.5
dLS
0
250
500
750
1000
1250
5
10
15
time

Thanks!

https://arxiv.org/pdf/2006.16901.pdf