Seminar “Statistics for structures”

A graphical perspective on Gauss-Markov process priors

Moritz Schauer
University of Amsterdam
Outline

- Midpoint displacement construction of a Brownian motion
- Corresponding Gaussian Markov random field
- Chordal graphs
- Sparse Cholesky decomposition
- Connection to inference of diffusion processes
Mid-point displacement

Lévy-Ciesielski construction of a Brownian motion \((W_t)_{t \in [0,1]}\)
Figure: Elements $\psi_{l,k}$, $1 \leq l \leq 3$ of the hierarchical (Faber-) Schauder basis
Schauder basis functions

A location and scale family based on the “hat” function
\[ \Lambda(x) = (2x)1_{[0, \frac{1}{2})} + 2(x - 1)1_{[\frac{1}{2}, 1]} \]

\[ \psi_{j,k}(x) = \Lambda(2^{j-1}x - k), \quad j \geq 1, \quad k = 0, \ldots, 2^{j-1} - 1 \]
Mid-point displacement II

Start with Brownian motion bridge \((W_t)_{t \in [0,1]}\)

\[
W^J = \sum_{j=1}^{J} \sum_{k=0}^{2j-1-1} Z_{j,k} \psi_{j,k}
\]

\(W^J\) – truncated Faber–Schauder expansion

\[
Z^J = \text{vec} \left( Z_{j,k}, j \leq J, 0 \leq k < 2^{j-1} \right)
\]

\(Z^J\) – independent zero mean Gaussian random variables

\[
Z_{j,k} = W_{2^{-j}(2k+1)} - \frac{1}{2} \left( W_{2^{-j+1}k} + W_{2^{-j+1}(k+1)} \right)
\]
Mid-point displacement II

Start with mean zero Gauss–Markov process \((W_t)_{t \in [0,1]}\)

\[
W^J = \sum_{j=1}^{J} \sum_{k=0}^{2^{j-1}-1} Z_{j,k} \psi_{j,k}
\]

\(W^J\) – truncated Faber–Schauder expansion

\[
Z^J = \text{vec} \left( Z_{j,k}, j \leq J, 0 \leq k < 2^{j-1} \right)
\]

\(Z^J\) – mean zero Gaussian vector with precision matrix \(\Gamma\)

\[
Z_{j,k} = W_{2^{-j}(2k+1)} - \frac{1}{2} (W_{2^{-j+1}k} + W_{2^{-j+1}(k+1)})
\]
Markov property

Write \( \iota := (j, k), \iota' = (j', k') \)

In general

\[ \Gamma_{\iota, \iota'} = 0 \quad \text{if} \quad Z_\iota \perp \perp Z_{\iota'} \mid Z_{\{\iota, \iota'\}^c} \]

By the Markov property

\[ \Gamma_{\iota, \iota'} = 0 \quad \text{if} \quad \psi_\iota \cdot \psi_{\iota'} \equiv 0 \]
Gaussian Markov random field

A Gaussian vector \((Z_1, \ldots, Z_n)\) together with the graph \(\mathcal{G}(\{1, \ldots, n\}, \mathcal{E})\) where

\[
\text{no edge in } \mathcal{E} \text{ between } \iota \text{ and } \iota' \quad \text{if} \quad Z_{\iota} \perp \perp Z_{\iota'} \mid Z_{\{\iota, \iota\}^c}
\]
A chordal graph is a graph in which all cycles of four or more vertices have a chord, which is an edge that is not part of the cycle but connects two vertices of the cycle.
Interval graph

The open supports of $\psi_{j,k}$ form an interval graph on pairs $(j, k)$. Interval graphs are chordal graphs.

In red a cycle of four vertices with a blue chord\(^1\)

\(^1\)An interval graph is the intersection graph of a family of intervals on the real line. Interval graphs are chordal graphs.
Sampling from the prior

- Sample $J$
- Compute factorization $SS' = \Gamma^J$
- Solve by backsubstitution

$$L'Z = WN$$

with $WN$ – standard white noise

Hence: How to find sparse factors?
Perfect elimination ordering

“A perfect elimination ordering in a graph is an ordering of the vertices of the graph such that, for each vertex \( v \), \( v \) and the neighbors of \( v \) that occur after \( v \) in the order form a clique.”

Example:

\[
(3, 0) \ (3, 1) \ (3, 2) \ (3, 4) \ (2, 0) \ (2, 1) \ (1, 0)
\]
Ordering the columns and rows of $\Gamma$ according to the perfect elimination ordering of the chordal graph: $\tilde{S}$ is the sparse Cholesky factor of $\tilde{\Gamma}$

$\tilde{\Gamma} = \begin{pmatrix} \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \\ \square & \square & \square & \square & \square & \square & \square \end{pmatrix}$

$\tilde{S} = \begin{pmatrix} \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \\ \square & \square & \square & \square \end{pmatrix}$

Cholesky decomposition has no fill in!
Exploiting hierarchical structure

Order rows and columns of $\Gamma$ according to the location of the maxima of $\psi_{j,k}$. $\Gamma$ has sparsity structure

$$(3, 0) (2, 0) (3, 1) (1, 0) (3, 2) (2, 1) (3, 3)$$

$$\Gamma = SS'$$ where

$$S = \begin{pmatrix}
\end{pmatrix}.$$
Recursive sparsity pattern

\[ S^1 = (s_{11}) \]

\[ S^J = \begin{bmatrix}
S_l^{J-1} & 0 & 0 \\
S_{cl} & s_{cc} & S_{cr} \\
0 & 0 & S_r^{J-1}
\end{bmatrix} \begin{bmatrix}
2^{J-1} - 1 \\
1 \\
2^{J-1} - 1
\end{bmatrix} \]
Hierarchical back-substitution

A hierarchical back-substitution problem of the form

\[
\begin{bmatrix}
S_l & 0 & 0 \\
S_{cl} & s_{cc} & S_{cr} \\
0 & 0 & S_r
\end{bmatrix}
\begin{bmatrix}
X_l \\
x_c \\
X_r
\end{bmatrix}
= \begin{bmatrix}
B_l \\
b_c \\
B_r
\end{bmatrix}
\]

\[(m+1+m) \times (m+1+m)\]

can be recursively solved by solving the back-substitution problems \(S_l X_l = B_l\), \(S_r X_r = B_r\) and setting

\[x_c = s_{cc}^{-1} \cdot (b_c - S_{cl} X_l - S_{cr} X_r)\]
Factorization in quasi linear time

\[
\begin{bmatrix}
A_l & A'_cl & 0 \\
A_{cl} & a_{cc} & A_{cr} \\
0 & A'_{cr} & A_r
\end{bmatrix}
= \begin{bmatrix}
S_l & 0 & 0 \\
S_{cl} & s_{cc} & S_{cr} \\
0 & 0 & S_r
\end{bmatrix}
\begin{bmatrix}
S_l & S_{cr} & 0 \\
0 & s_{cc} & 0 \\
0 & S_{cr} & S_r
\end{bmatrix}
\]

\[
= \begin{bmatrix}
S_lS'_l \\
S'_{cl}S_l \\
0
\end{bmatrix}
\begin{bmatrix}
S'_lS_{cl} \\
s_{cc}^2 + S_{cl}S'_{cl} + S_{cr}S'_{cr} \\
0
\end{bmatrix}
\begin{bmatrix}
S'_lS_{cr} \\
S'_{cr}S_r \\
S'_rS_r
\end{bmatrix}
\]

Here \( A_l = S_lS'_l \) and \( A_r = S_rS'_r \) are two hierarchical factorization problems of level \( J - 1 \), \( A_l = S'_{cl}S_l \) and \( A_r = S'_{cr}S_r \) are hierarchical back-substitution problems and

\[
s_{cc} = \sqrt{a_{cc} - S_{cl}S'_{cl} + S_{cr}S'_{cr}}.
\]
Approximative sparse inversion using nested dissection
Application: Nonparametric inference for diffusion process

\[ dX_t = b_0(X_t)\, dt + dW_t \]  

(1)

Prior \( P(J \geq j) \geq C \exp(-2^j) \) and

\[
b = \sum_{j=1}^{J} \sum_{k=0}^{2^{j-1}-1} Z_{j,k} \psi_{j,k}
\]

where \( \alpha = \frac{1}{2} \), \( \Xi^J = \text{diag}(2^{-2(j-1)\alpha}, 1 \leq j \leq J, 0 \leq k < 2^{j-1}) \)
Gaussian inverse problem

Likelihood

\[ p(X \mid b) = \exp \left( \int_0^T b(X_t) \, dX_t - \frac{1}{2} \int_0^T b^2(X_t) \, dt \right) \]

\[ \mu^J_\iota = \int_0^T \psi_\iota(X_t) \, dX_t, \quad \iota = 1, \ldots, 2^J - 1 \]

\[ G^J_{\iota,\iota'} = \int_0^T \psi_\iota(X_t) \psi_{\iota'}(X_t) \, dt, \quad \iota, \iota' = 1, \ldots, 2^J - 1. \]

\( \Gamma^J \) and \( G^J \) have the same sparsity pattern
Conjugate posterior

For fix level $J$,

$$Z^J \mid J, X \sim \mathcal{N}(\Sigma^J \mu^J, \Sigma^J)$$

where $\Sigma^J = (\Gamma^J + G^J)^{-1}$. 

On $J$ a reversible jump algorithm can be used.
Posterior contraction rates (periodic case)

Besov norm, supremum norm for $f = \sum \sum z_{j,k} \psi_{j,k}$

$$\| f \|_\alpha = \sup_{j \geq 1, k} 2^{(j-1)\alpha} |z_{j,k}| \quad \| f \|_\infty \leq \sum_j \max_k |z_{j,k}|$$

Sieves

$$B_{L,M} = \left\{ \sum_{j=1}^{L} \sum_{k=0}^{2^{j-1}-1} z_{j,k} \psi_{j,k} : 2^{\alpha(j-1)} |z_{j,k}| \leq M, j, k = \ldots \right\}$$

Rate

$$T^{-\frac{\beta}{1+2\beta}} \log(T)^{\frac{\beta}{1+2\beta}} \quad \beta \geq \alpha$$
Anderson’s lemma

If $X \sim N(0, \Sigma_X)$ and $Y \sim N(0, \Sigma_Y)$ independent with $\Sigma_X \preceq_{pd} \Sigma_Y$ positive definite, then for all symmetric convex sets $P(Y \in C) \leq P(X \in C)$. 
Summary

- Midpoint displacement construction of Gauss-Markov processes
- Corresponding Gaussian Markov random field
- Chordal graphs and perfect elimination orderings
- Sparse Cholesky decomposition
- Rates for randomly truncated prior
Image sources