**Hierarchical Likelihood Approximation**

**Goal:** To improve estimation of unknown statistical parameters in a spatial soil moisture field.

**How?** By reducing linear algebra cost from $O(n^3)$ to $O(n \log n)$.

Let $Z$ be a mean-zero, stationary and isotropic Gaussian process with a Matérn covariance at $n$ irregularly spaced locations.

Let $Z = (Z(s_1), \ldots, Z(s_n))^T \sim N(0, C(\theta))$, $\theta \in \mathbb{R}^d$ is an unknown parameter vector of interest, where

$$C_{ij}(\theta) = \text{cov}(Z(s_i), Z(s_j)) = C(|s_i - s_j|, \theta),$$ and

$$C(r) := C_\theta(r) = \frac{2^{\nu-1}}{\Gamma(\nu)} \left( \frac{\nu r}{\ell} \right)^\nu \frac{K_\nu \left( \frac{\nu r}{\ell} \right)}{\left( \frac{\nu r}{\ell} \right)^\nu}, \quad \theta = (\sigma^2, \nu, \ell)^T,$

is the Matérn covariance function. The MLE of $\theta$ is obtained by maximizing the Gaussian log-likelihood function:

$$L(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(C(\theta)) - \frac{1}{2} C^{-1} Z\theta.$$

We approximate $C \approx \hat{C}$ in the $H$-matrix format with cost and storage $O(kn \log n)$, $k \ll n$. Obtain a cheap approximation $L(\hat{\theta}) \approx \hat{L}(\theta, k)$.

A realization of a Matérn random field

Simulated data with known parameters $(\ell^*, \nu^*, \sigma^*) = (0.0864, 0.5, 1.0)$. Boxplots for $\ell$ and $\nu$ for $n = \{128, 64, 32, 16, 8, 4, 2\} \times 1000$. Boxplots are obtained from 50 replicates, $\varepsilon = 10^{-9}$.

**Numerical Examples**

$H$-matrix approximation, $\nu = 0.5$, domain $\mathcal{G} = [0, 1]^2$, $\|\hat{C}(0.25, 0.75)\|_2 = (212.568)$, $n = 16049$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>KLD $\ell = 0.25$</th>
<th>$|C - \hat{C}|_2$, $\ell = 0.25$</th>
<th>$|CC^{-1} - I|_2$, $\ell = 0.25$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.6e-3, 0.2</td>
<td>7.6e-4, 7.6e-4</td>
<td>6.0e-2, 3.1</td>
</tr>
<tr>
<td>50</td>
<td>3.4e-13, 5e-12</td>
<td>2.0e-13, 2.4e-13</td>
<td>4e-11, 2.7e-9</td>
</tr>
</tbody>
</table>

Computing time and number of iterations for maximization of log-likelihood $\hat{L}(\theta, k)$, $n = 66049$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>size, GB</th>
<th>$C$, set up time, s.</th>
<th>compute $L$, s.</th>
<th>maximizing, s.</th>
<th># iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1</td>
<td>7</td>
<td>115</td>
<td>1994</td>
<td>13</td>
</tr>
<tr>
<td>20</td>
<td>1.7</td>
<td>11</td>
<td>370</td>
<td>5445</td>
<td>9</td>
</tr>
<tr>
<td>dense</td>
<td>38</td>
<td>42</td>
<td>657</td>
<td>$\infty$</td>
<td>-</td>
</tr>
</tbody>
</table>

**Parallel Hierarchical Matrices (Hackbusch, Kriemann’05)**

Advantages to approximate $C$ by $\hat{C}$: $H$-approximation is cheap; storage and matrix-vector product cost $O(kn \log n)$; LU and inverse cost $O(k^2 n \log n)$; efficient parallel implementations exists.

(1st) Matérn $H$-matrix approximations for moisture example, $n = 8000$, $\varepsilon = 10^{-3}$, $\ell = 0.64$, $\nu = 0.325$, $\sigma^2 = 0.98$, 29.3MB vs 488.3MB for dense, set up time 0.4 sec.; (2nd) $H$-Cholesky factor $\hat{L}$, with accuracy in each block $\varepsilon = 10^{-8}$, 4.8 sec., storage 52.8 MB; (3rd) Dependence of the negative log-likelihood $-\hat{L} / n$, on $n = \{2000, \ldots, 128000\}$ and on the parameters $\ell$ and $\nu$ (4th), in log-log scale for the soil moisture data example; (5th) error ($\|I - (\hat{L} \hat{L}^{-1}) L\|_2$) vs. $\nu$ ($\hat{\nu} = 0.325$, $\hat{\sigma}^2 = 0.98$ fixed), $\varepsilon = 10^{-4}$.

**References and Acknowledgements**


Work supported by SRI-UQ and ECRC, KAUST. Thanks to Ronald Kriemann for HLlibPro.