Hierarchical matrix approximation of large covariance matrices

A. Litvinenko, M. Genton, Y. Sun, SRI-UQ Center, Spatio-Temporal Statistics Group and Extreme Computing Research Center at King Abdullah University of Science and Technology (KAUST)

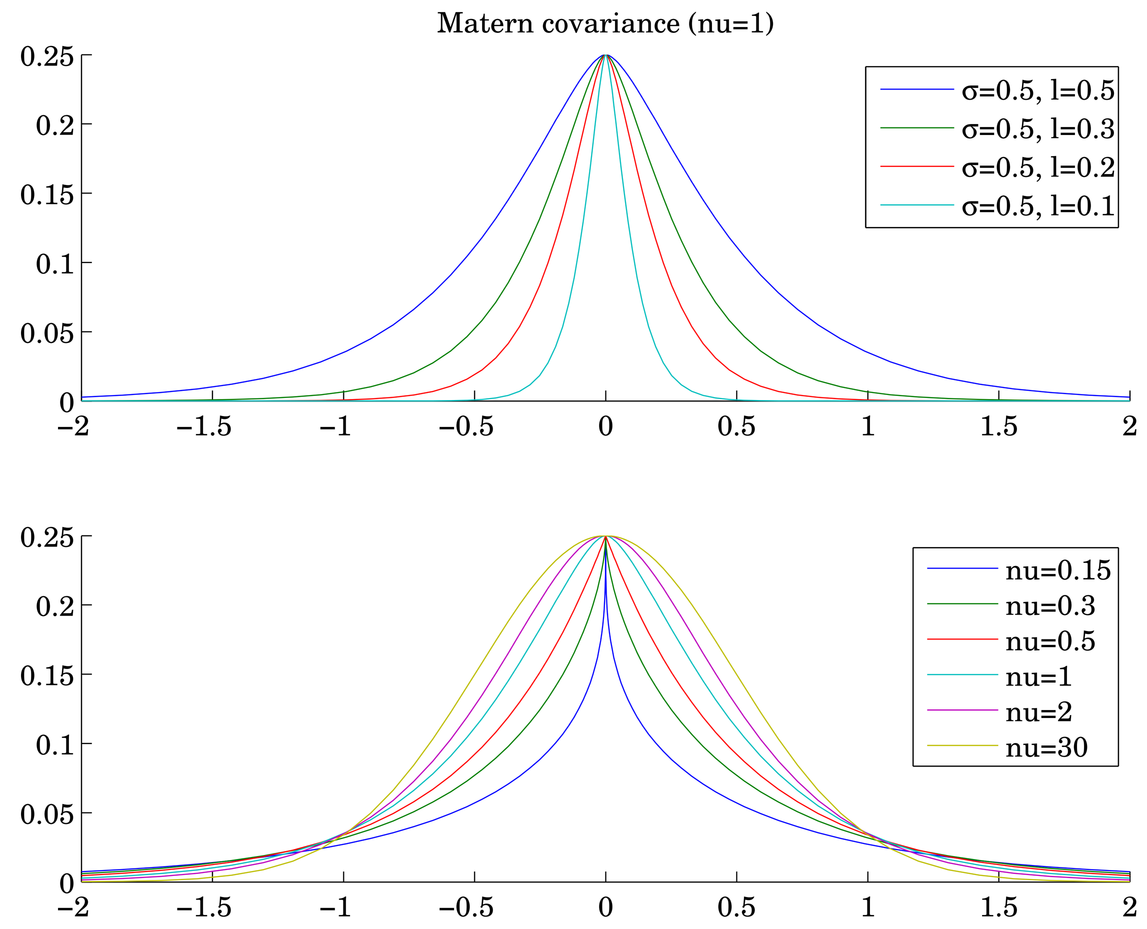
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Maximum likelihood covariance estimation

We approximate large non-structured Matérn covariance matrices of size $n\times n$ in the $\mathcal{H}$-matrix format with a log-linear computational cost and storage $\mathcal{O}(kn\log n)$, where rank $k \ll n$ is a small integer. Applications are: spatial statistics, machine learning and image analysis, kriging and optimal design.

Matérn covariance function is defined as

$$C(r) := C_\theta(r) = \frac{2\sigma^2}{\Gamma(\nu)}\left(\frac{r}{2l}\right)^\nu K_\nu\left(\frac{r}{l}\right),\ \theta = (\sigma^2, \nu, l).$$



Shape of Matérn covariance functions.

Suppose we observe $\boldsymbol{z} = (z_1, ..., z_n)^T \sim \mathcal{N}(0, C(\theta^*))$ for some fixed $\theta = \theta^*$ (truth to be identified). To estimate uncertain parameter θ we maximize the log-likelihood

$$-2\ell(\theta) = N\log 2\pi + \log\det\{C(\theta)\} + \boldsymbol{z}^T C(\theta)^{-1}\boldsymbol{z},$$

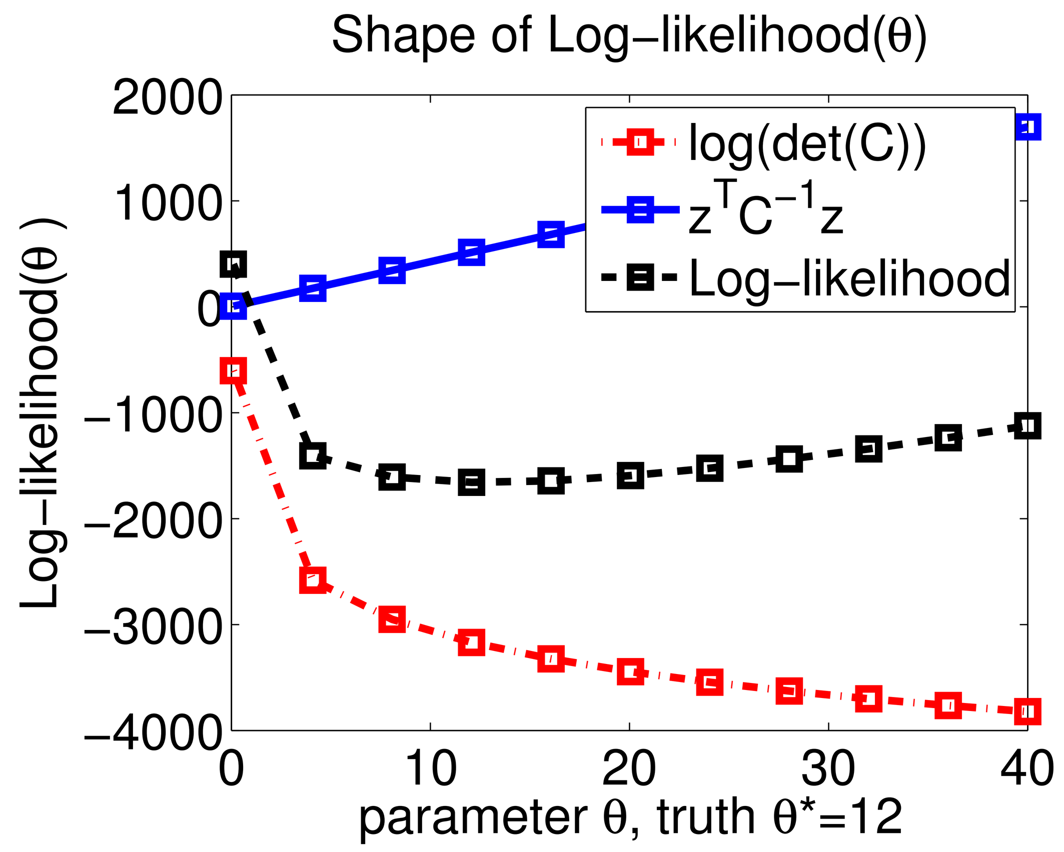
where $C = LL^T$, $\boldsymbol{z} = L(\theta^*)\cdot y$, $y \sim \mathcal{N}(0, I)$- normal distributed. On each iteration i of a maximization algorithm we have a new matrix $C(\theta_i)$. Approximate the likelihood $\ell(\theta)$ by $\tilde{\ell}(\theta, k)$, where k is the $\mathcal{H}$-matrix rank used for approximation of the covariance matrix $C(\theta) \approx C^{\mathcal{H}}(\theta, k)$.

Maximize $\tilde{\ell}(\theta, k)$ and find the parameter $\hat{\theta}(k)$, where this maximum is achieved. Here $C(\theta)^{-1}\boldsymbol{z}(\theta^*)$ is computed in two steps

$$1) L(\theta)x(\theta) = \boldsymbol{z}(\theta^*) \quad\text{and}\quad 2) L(\theta)^T r(\theta) = x(\theta).$$

$$\log\det\{C\} = \log\det\{\prod_{i=1}^n \lambda_i^2\} = 2\sum_{i=1}^n \log\lambda_i,$$

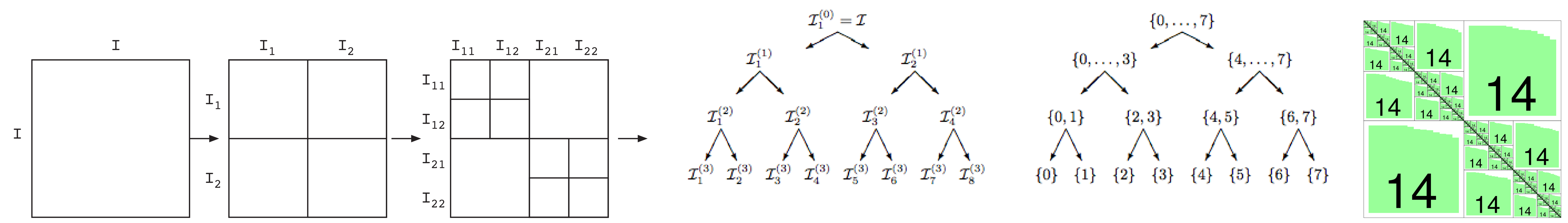
where λ_i are diagonal elements of $\mathcal{H}$-Cholesky factor L.



Minimum of negative log-likelihood (black) is at $\theta = (\cdot,\cdot,l) \approx 12$ (σ^2 and ν are fixed).

Hierarchical matrices

Advantages to approximate C by $C^{\mathcal{H}}$: $\mathcal{H}$-approximation is cheap, storage and MV $\mathcal{O}(kn\log n)$, LU and M^{-1} $\mathcal{O}(k^2 n\log^2 n)$, efficient parallel implementations exist.



$\mathcal{H}$-matrix approximations of exponential covariance $C \approx C^{\mathcal{H}} \in \mathbb{R}^{n\times n}$, $n = 32^2$, cov. lengths in x and in y directions are 0.15 and 0.2, domain $[0,1]^2$. Dense (red/dark) blocks $\in \mathbb{R}^{32\times 32}$, max. rank $k = 14$.

Numerical examples

n	rank k for $C^{\mathcal{H}}$	size, MB C	size, MB $C^{\mathcal{H}}$	t, sec. C	t, sec. $C^{\mathcal{H}}$	ε	$\max_{i=1..10} \lvert\lambda_i - \tilde{\lambda}_i\rvert$, i	ε_2
$4.0\cdot 10^3$	10	48	3	0.8	0.08	$7\cdot 10^{-3}$	$7.0\cdot 10^{-2}$, 9	$2.0\cdot 10^{-4}$
$1.05\cdot 10^4$	18	439	19	7.0	0.4	$7\cdot 10^{-4}$	$5.5\cdot 10^{-2}$, 2	$1.0\cdot 10^{-4}$
$2.1\cdot 10^4$	25	2054	64	45.0	1.4	$1\cdot 10^{-5}$	$5.0\cdot 10^{-2}$, 9	$4.4\cdot 10^{-6}$

$\mathcal{H}$-matrix approximation of exp. cov. function, cov. lengths in x, y, z directions are 0.1, 0.5 and 0.1. Column 6. Accuracy of 10 largest eigenvalues. $\varepsilon_2 := \frac{\lvert\|C\|_2 - \|\tilde{C}\|_2\rvert}{\|C\|_2}$, $\varepsilon := \frac{\|(C-\tilde{C})v\|_2}{\|C\|_2\|v\|_2}$, v - random vector.

Kullback-Leibler divergence (KLD) for miltivariate normal distributions (μ_0, C) and $(\mu_1, C^{\mathcal{H}})$ is

$$D_{KL}(\mathcal{N}_0\|\mathcal{N}_1) = 0.5\left(tr((C^{\mathcal{H}})^{-1}C) + (\mu_1-\mu_0)^T(C^{\mathcal{H}})^{-1}(\mu_1-\mu_0) - n - \log_e\left(\frac{\det C}{\det C^{\mathcal{H}}}\right)\right)$$

k	KLD $l=0.25$	KLD $l=0.75$	$\|C-C^{\mathcal{H}}\|_2$ $l=0.25$	$\|C-C^{\mathcal{H}}\|_2$ $l=0.75$	$\|C(C^{\mathcal{H}})^{-1}-I\|_2$ $l=0.25$	$\|C(C^{\mathcal{H}})^{-1}-I\|_2$ $l=0.75$
10	2.6e-3	0.2	7.7e-4	7.0e-4	6.0e-2	3.1
15	1.0e-5	9e-4	2.0e-5	1.1e-5	8.0e-4	0.02
50	3.4e-13	5e-12	2.0e-13	2.4e-13	4e-11	2.7e-9

Dependence of KLD on the approximation $\mathcal{H}$-matrix rank k, Matérn covariance with cov. length $\{0.25, 0.75\}$ in x and y directions and $\nu = 0.5$, domain $\mathcal{G} = [0,1]^2$, $\|C_{(0.25,0.75)}\|_2 = \{212, 568\}$, $n = 129^2$.

matrix, $n\times n$	n	size, GB	set up t, s.	$\mathcal{H}$-Chol. t, s.	t maximizing, s.	# iters
$\mathcal{H}$-matrix, $k=10$	66049	1	7	115	1994	13
$\mathcal{H}$-matrix, $k=20$	66049	1.7	11	370	5445	9
dense	66049	38	42	657	∞	-

Detailed computing time and number of iterations for maximization of log-likelihood $\tilde{\ell}(\theta, k)$.

Conclusion

1. We have successfully applied $\mathcal{H}$-matrix technique for the approximation of covariance matrices.
2. Parallel $\mathcal{H}$-matrix approximation is natural for huge covariance matrices.
3. Further low-rank approximation of a) Kriging, b) variance and c) geostatistical optimal design:

a) Kriging estimate $s = C_{sy}C_{yy}^{-1}y$, y - measurements, C_{sy} and C_{yy} are cross and auto-covariance matrices.
b) $\sigma = \mathrm{diag}(C_{ss|y}) = \mathrm{diag}(C_{ss} - C_{sy}C_{yy}^{-1}C_{ys})$
c) $\phi_A = n^{-1}\mathrm{trace}(C_{ss|y})$, $\phi_C = v^T\left(C_{ss} - C_{sy}C_{yy}^{-1}C_{ys}\right)v$, v a vector.

References and Acknowledgements

[1] B. N. Khoromskij, A. Litvinenko, H. G. Matthies, *Application of hierarchical matrices for computing the Karhunen-Loe?ve expan-sion*, Computing, Vol. 84, Issue 1-2, pp 49-67, 2008.

[2] Y. Sun, M. Stein, *Statistically and computationally efficient estimating equations for large spatial datasets*, JCGS,? 2015, to appear.

[3] J. Castrillon-Candas, M. Genton, and R. Yokota, *Multi-Level Restricted Maximum Likelihood Covariance Estimation and Kriging for Large Non-Gridded Spatial Datasets*, arXiv:1504.00302v1, 2015.

[4] W. Nowak, A. Litvinenko, *Kriging and spatial design accelerated by orders of magnitude: combining low-rank covariance approxi-mations with FFT-techniques*, JMathematical Geosciences, Vol. 45, N4, pp411-435, 2013.

Work supported by SRI-UQ and ECRC, KAUST.