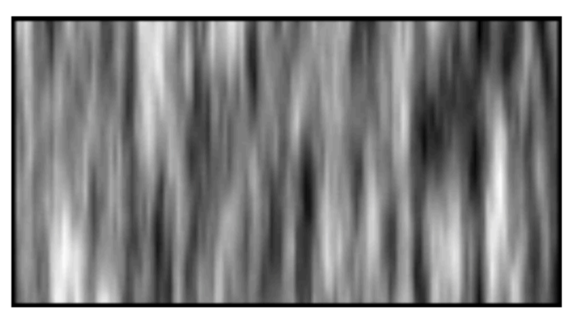
Fast Spatial Gaussian Process Maximum Likelihood Estimation via Skeletonization Factorizations

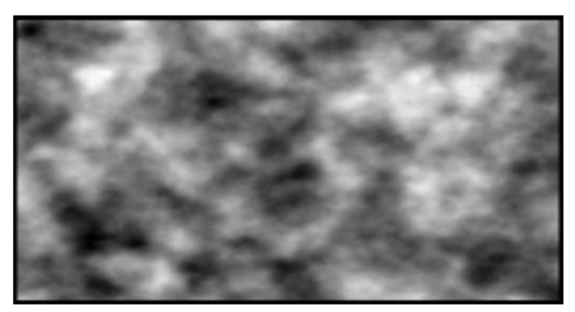
Victor Minden, Anil Damle, Ken Ho, Lexing Ying

Institute for Computational & Mathematical Engineering / Department of Mathematics, Stanford University

Introduction

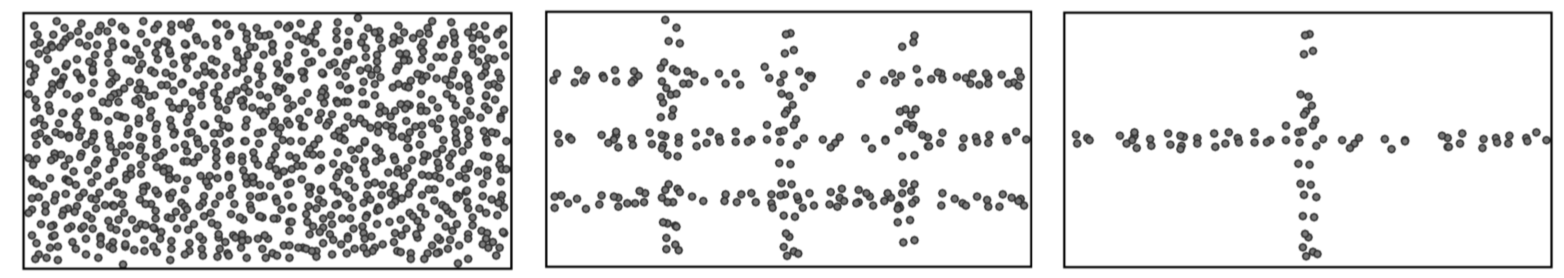
- Gaussian processes (Kriging) used frequently to model spatial fields in 2D.
- For data sets with many observations, limited by *O(N³)* cost of working with large kernelized covariance matrices.
- Using hierarchical matrix structure, we show a fast way to compute maximum likelihood estimates for parameterized kernels.





$$k_M(x,y;\theta) = (1+\sqrt{3}\|x-y\|_\theta)\exp(-\sqrt{3}\|x-y\|_\theta) + \sigma^2 I$$

Numerical Results

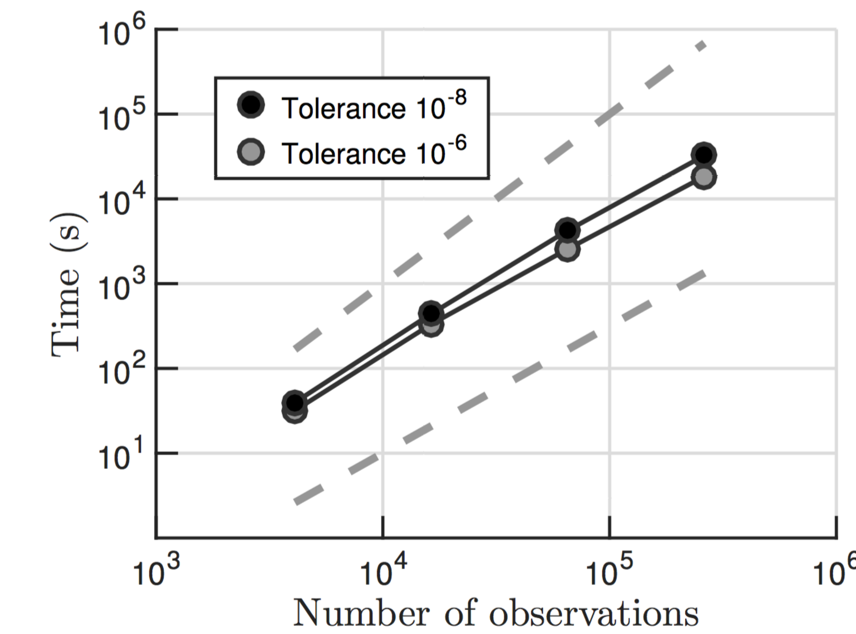


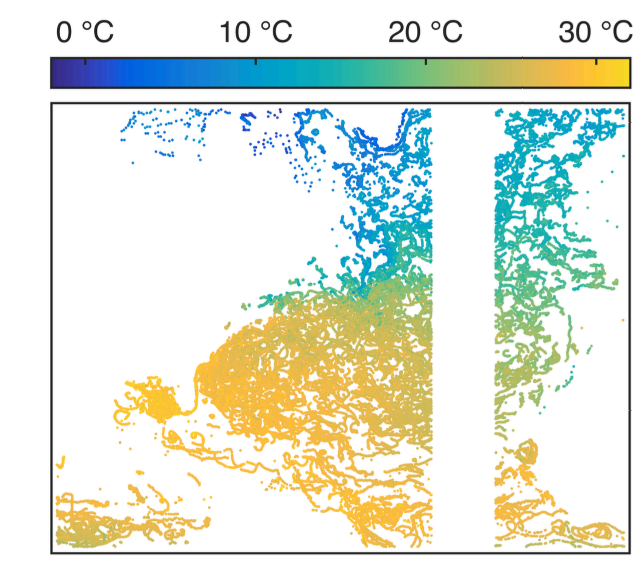
Algorithm 4.1 Computing the Gaussian process log-likelihood and gradient

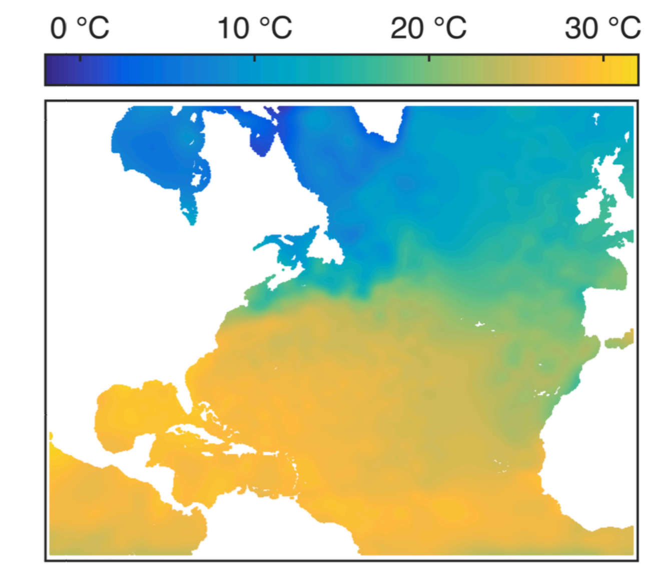
Given: observation vector $z \in \mathbb{R}^n$, parameter vector $\theta \in \mathbb{R}^p$, peel tolerance ϵ_{peel}, factorization tolerance $\epsilon_{\text{fact}} < \epsilon_{\text{peel}}$ and covariance kernel family $k(\cdot,\cdot;\theta)$

`// Factor` Σ `with hierarchial factorization`
$F \leftarrow$ Recursive skeletonization factorization of Σ with tolerance ϵ_{fact}
`// Use fast hierarchical solve and log-determinant`
$\hat{\ell}(\theta) \leftarrow -\frac{1}{2}z^T F^{-1} z - \frac{1}{2}\log|F| - \frac{1}{2}\log 2\pi \approx \ell(\theta)$
for $i = 1,\dots,p$ **do**
    `// Factor` Σ_i `with hierarchical factorization`
    $F_i \leftarrow$ Recursive skeletonization factorization of Σ_i with tolerance ϵ_{fact})
    `// Compute trace of` $\Sigma^{-1}\Sigma_i$ `with peeling algorithm`
    $t_i \leftarrow$ Trace of peeled operator $\frac{1}{2}(F^{-1}F_i + F_i F^{-1})$ via peeling algorithm with tolerance ϵ_{peel}
    `// Use fast hierarchical apply and solve`
    $\hat{g}_i \leftarrow \frac{1}{2}z^T F^{-1} F_i F^{-1} z - \frac{1}{2}t_i \approx g_i$
end for

Output: $\hat{\ell}(\theta)$ and $\hat{g}$



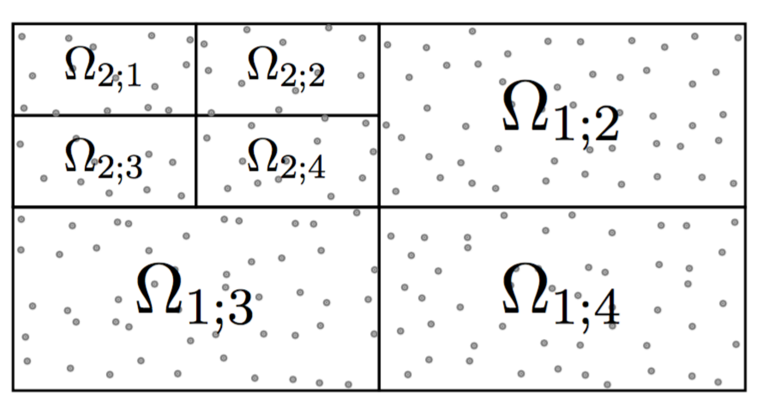




Top: runtime scaling for one iteration of maximum likelihood estimation with our method.
Right: Kriged estimates of sea-surface temperature in the Atlantic ocean using restricted maximum likelihood estimation on a subset of the ICOADS data set.

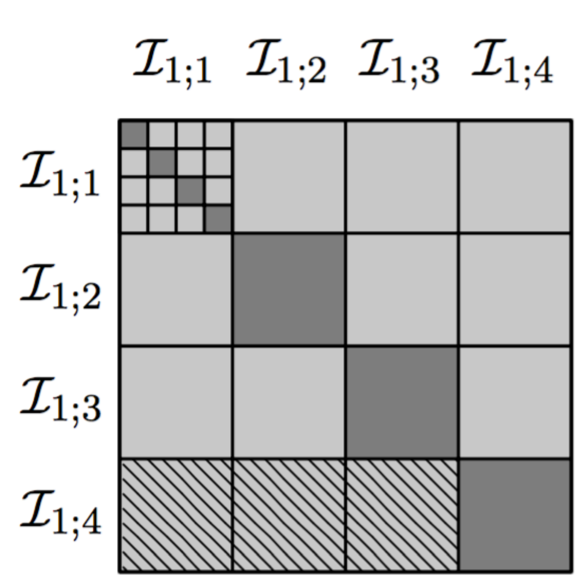
Basic Idea

- Log-likelihood is
$$\ell(\theta) = -\frac{1}{2}z^T\Sigma^{-1}z - \frac{1}{2}\log|\Sigma| - \frac{n}{2}\log 2\pi$$
with gradient
$$g_i = \frac{\partial \ell(\theta)}{\partial \theta_i} = \frac{1}{2}z^T\Sigma^{-1}\Sigma_i\Sigma^{-1}z - \frac{1}{2}\mathrm{Tr}(\Sigma^{-1}\Sigma_i)$$



- Use recursive skeletonization factorization to approximate
$$\Sigma \approx \left[\prod_{\ell=1}^{L}\left(\prod_{\mathcal{I}\in\mathscr{L}_\ell} L_{\mathcal{I}}U_{\mathcal{I}}\right)P_\ell\right]\tilde{\Sigma}_0\left[\prod_{\ell=1}^{L}\left(\prod_{\mathcal{I}\in\mathscr{L}_\ell} L_{\mathcal{I}}U_{\mathcal{I}}\right)P_\ell\right]^T$$

- Use black-box applies to construct H-matrix to "peel" trace term.



Rank r_ℓ of off-diagonal blocks of G at level ℓ	Time	Storage
$O(1) - O(\log n_\ell)$	$\tilde{O}(n)$	$\tilde{O}(n)$
$O(\sqrt{n_\ell})$	$\tilde{O}\left(n^2\right)$	$\tilde{O}(n^{3/2})$

Key References

S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O'Neil. Fast direct methods for Gaussian processes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(2):252–265, Feb 2016.

J. E. Castrillón-Candás, M. G. Genton, and R. Yokota. Multi-level restricted maximum likelihood covariance estimation and Kriging for large non-gridded spatial datasets. Spatial Statistics, 2015.

W. Hackbusch, N. B. Khoromskij, and R. Kriemann. Hierarchical matrices based on a weak admissibility criterion. Computing, 73(3):207–243, 2004.

K. L. Ho and L. Greengard. A fast direct solver for structured linear systems by recursive skeletonization. SIAM Journal on Scientific Computing, 34(5):A2507–A2532, 2012.

K. L. Ho and L. Ying. Hierarchical interpolative factorization for elliptic operators: Integral equations. Communications on Pure and Applied Mathematics, 2015.

L. Lin, J. Lu, and L. Ying. Fast construction of hierarchical matrix representation from matrix-vector multiplication. Journal of Computational Physics, 230(10):4071–4087, 2011.

P. G. Martinsson and V. Rokhlin. A fast direct solver for boundary integral equations in two dimensions. Journal of Computational Physics, 205(1):1–23, May 2005.

M. L. Stein. Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics. Springer New York, 1999. ISBN 9780387986296.

Full references in paper (see *http://arxiv.org/abs/1603.08057* or QR code)

Acknowledgments

V.M. is supported by a U.S. Department of Energy Computational Science Graduate Fellowship under grant number DE-FG02-97ER25308. A.D. is partially supported by a National Science Foundation Graduate Research Fellowship under grant number DGE-1147470 and a Simons Graduate Research Assistantship. K.H. is supported by a National Science Foundation Mathematical Sciences Postdoctoral Research Fellowship under grant number DMS-1203554. L.Y. is supported by the National Science Foundation under award DMS-1328230 and DMS-1521830 and the U.S. Department of Energy's Advanced Scientific Computing Research program under award DE-FC02-13ER26134/DE-SC0009409.