Gaussian Process based models provide much inferential power, as well as variance information, however they require the solution of dense linear systems with size proportional to the number of points, making the direct approach infeasible for large problems. This talk discusses the use of preconditioners for hierarchical matrix based solutions to accelerate Gaussian Process prediction and optimization. The use of the SMASH H^2 hierarchical matrix from [2] results in near linear matrix-vector products, which can be used in iterative solvers. We consider and present results for a variety preconditioners including both geometry aware and global preconditioners.

Given a symmetric positive definite kernel $K: K(\cdot, \cdot) \in \mathbb{R}$ with training points $X$, test points $X_*$, and training and test output vectors $f$ and $f_*$ respectively, we have the multivariate Gaussian distribution of the output

$$
\begin{bmatrix}
  f \\
  f_*
\end{bmatrix} \sim \mathcal{N}
\begin{pmatrix}
  0, \\
  K(X, X) & K(X, X_*)
\end{pmatrix}
$$

This results in the predictive distribution

$$
f_*|X, f, X_* \sim \mathcal{N}(\text{mean, covariance})$$

mean : $K(X_*, X)(K(X, X))^{-1} y$ (2)

covariance : $K(X_*, X) - K(X_*, X)(K(X, X))^{-1} K(X, X_*)$ (3)

In Gaussian Process literature, the direct way of performing a Gaussian Process prediction involves computing the solutions of the linear system via the Cholesky decomposition of the $K(X, X)$ covariance matrix, an $O(n^3)$ operation [5]. This $O(n^3)$ complexity makes Gaussian Process regression prohibitively expensive for large problems, and as such there is great motivation to develop methods which decrease the complexity. The most popular kernel is the Gaussian Radial Basis, or Squared Exponential, kernel

$$
K(x, y) = e^{-\frac{1}{2} \frac{|x-y|^2}{\sigma^2}},
$$

---

1 Georgia Institute of Technology
2 Georgia Institute of Technology
3 Purdue University
4 Emory University
5 Georgia Institute of Technology
6 Georgia Institute of Technology
which is smooth, and thus the covariance matrix is a good candidate for representation via a hierarchical matrix.

The underlying idea of hierarchical matrices is that, for data sparse matrices, the matrix can be approximated via a hierarchy of matrices with a low-rank representation, and using properties of this approximation operations can be greatly accelerated. We use the SMASH $H^2$ matrix which provides parallelizable, near-linear matrix-vector products which can be used in matrix-free methods. While using hierarchical matrices for Gaussian Process Regression [1] and Kernel Ridge Regression [4] have been investigated previously, this talk focuses on the use of preconditioners in conjunction with iterative solvers accelerated by hierarchical matrices.

The need for preconditioners is due to the ill conditioning of the covariance matrix. One of the steps of the SMASH matrix construction is the construction of a cluster tree, which provides information on the geometry of the problem, which we able to exploit in our geometry aware preconditioners. We investigate the use of LU, SVD, and SVD with dynamic stabilization inside of this geometry aware framework. In addition to these preconditioners we also investigate and present results for Chebyshev polynomial preconditioners of varying degrees.

References


