Sparse Inverse Gaussian Process Regression and its Application to Climate Network Discovery

Kamalika Das[1] Ashok N. Srivastava [2]

[1]SGT Inc.—NASA Ames Research Center, Moffett Field CA 94035

[2]NASA Ames Research Center, Moffett Field CA 94035

CIDU 2011

Roadmap

Introduction

- Desired characteristics in prediction model: accuracy, confidence, scalability, interpretability
- Gaussian Process regression
  - Predicts a distribution (mean and variance)
  - Captures non-linear relationship in data
- Unknown sparsity structure
  - Calculating complete inverse covariance matrix will give much denser matrix
  - Low rank approximation techniques address scalability but destroy model interpretability
  - Can reveal important causal relationships in data

Contribution

- Sparse Gaussian Process regression using estimation of a sparse inverse covariance matrix
- Can be parallelized for scaling to large data sets
- Illustrative application on a climate domain data set

Background: Gaussian Process regression

- Training data
  - X: data matrix of observations $n \times d$
  - y: vector of target data $n \times 1$
- Test data
  - X^*: matrix of new observations $n^* \times d$
- Covariance function: $K_{ij} = k(x_i, x_j), K^*_{ij} = k(x^*_i, x_j)$

Prediction equation

$$\widehat{y}^* = K^*(\lambda^2 I + K)^{-1} y$$

Inverse covariance estimation

- Multivariate Gaussian distribution: $\mathbf{x}_i \sim \mathcal{N}(\mu, \Sigma)$
  - $\Sigma_{i,j} = 0 \Rightarrow P(\mathbf{x}_i \mathbf{x}_j) = 0$ and $\Sigma^{-1}_{i,j} = 0 \Rightarrow P(\mathbf{x}_i \mathbf{x}_j | \mathbf{x}_{-i,-j}) = 0$

- Gaussian Process
  - Defines a distribution over functions, specified by mean function and covariance function

$$f(\mathbf{x}_i) \sim GP(m(\mathbf{x}_i), k(\mathbf{x}_i, \mathbf{x}_j))$$

  - $m(\mathbf{x}_i) = E[f(\mathbf{x}_i)]$ and $k(\mathbf{x}_i, \mathbf{x}_j) = E[f(\mathbf{x}_i) - m(\mathbf{x}_i)][f(\mathbf{x}_j) - m(\mathbf{x}_j)]$

Inverse covariance estimation

- Multivariate Gaussian distribution: $\mathbf{x}_i \sim \mathcal{N}(\mu, \Sigma)$
  - $\Sigma_{i,j} = 0 \Rightarrow P(\mathbf{x}_i \mathbf{x}_j) = 0$ and $\Sigma^{-1}_{i,j} = 0 \Rightarrow P(\mathbf{x}_i \mathbf{x}_j | \mathbf{x}_{-i,-j}) = 0$

- Gaussian Process
  - Defines a distribution over functions, specified by mean function and covariance function

$$f(\mathbf{x}_i) \sim GP(m(\mathbf{x}_i), k(\mathbf{x}_i, \mathbf{x}_j))$$

  - $m(\mathbf{x}_i) = E[f(\mathbf{x}_i)]$ and $k(\mathbf{x}_i, \mathbf{x}_j) = E[f(\mathbf{x}_i) - m(\mathbf{x}_i)][f(\mathbf{x}_j) - m(\mathbf{x}_j)]$

Theorem

Covariance selection for graphical models is equivalent to inverse covariance matrix estimation in Gaussian Process

Inverse covariance estimation

- Inverse covariance estimation gives relevant conditional independence information

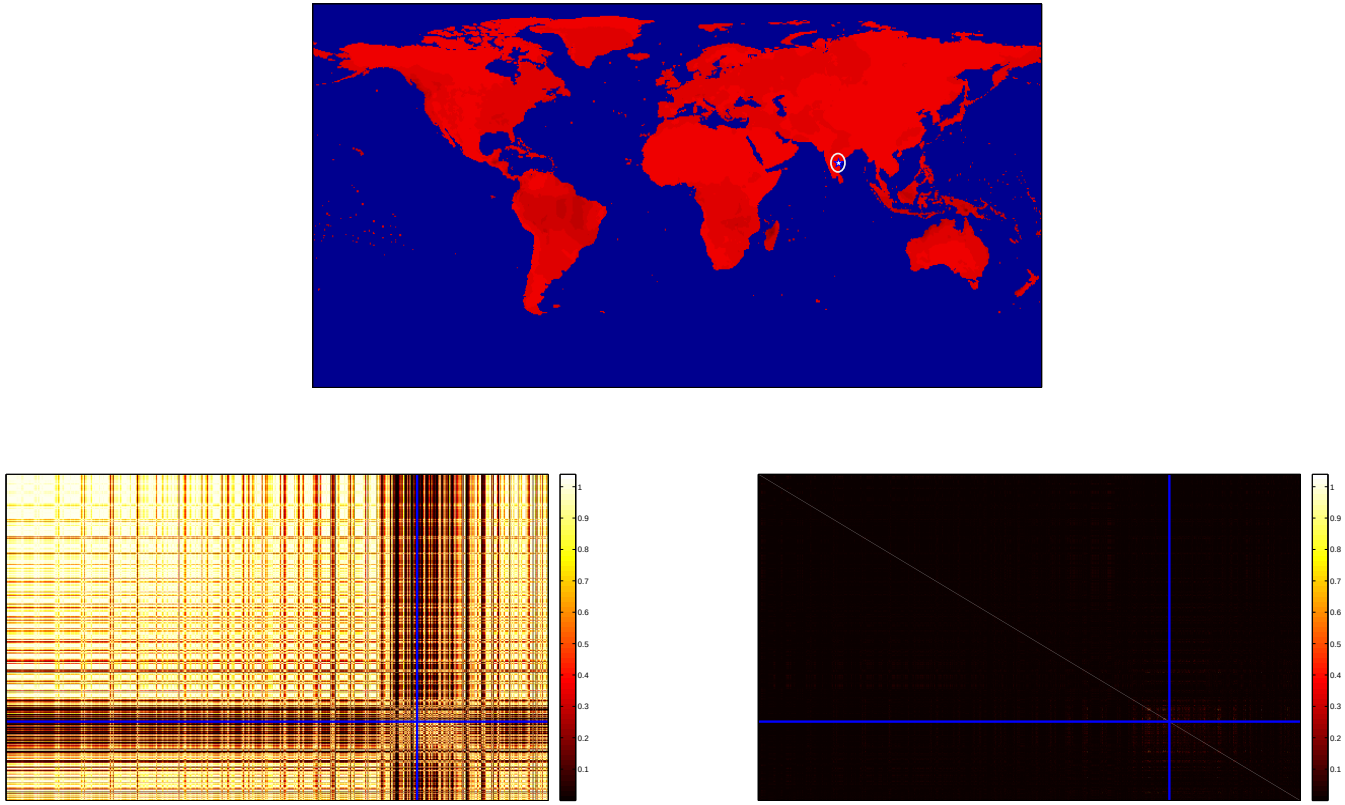


Figure: Sample location in India, kernel and inverse kernel matrix

Sparse inverse covariance estimation

- Minimize the pseudo negative log likelihood

Maximum likelihood estimation

$$\mathbf{Tr}(KS) - \log det(S)$$

Sparse inverse covariance estimation

- Minimize the pseudo negative log likelihood

Maximum likelihood estimation

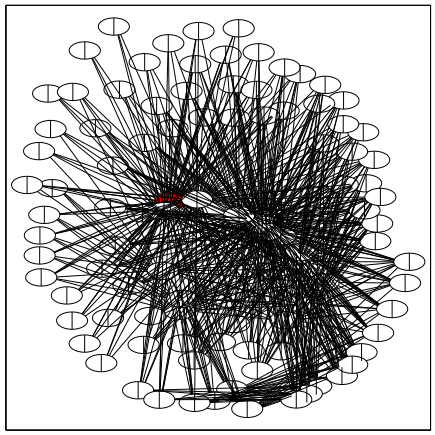
$$\mathbf{Tr}(KS) - \log det(S)$$

- Induce sparsity using the ℓ_1 regularizer

Sparse maximum likelihood estimation

$$\mathbf{Tr}(KS) - \log det(S) + \lambda \|S\|$$

Sparse inverse covariance estimation



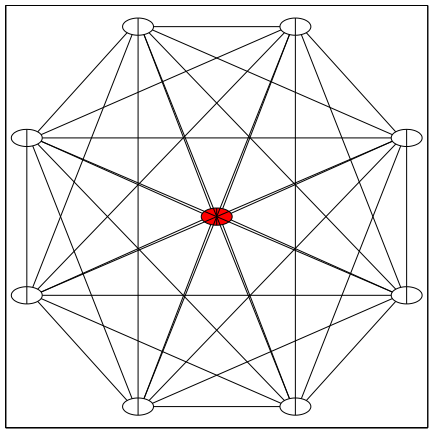


Figure: Inverse (left) and sparse inverse (right)

Alternating direction method of multipliers

- Decomposition algorithm for solving convex optimization problems
- Based on iterative scatter and gather operations on the augmented Lagrangian
- Solves problems of the form:

$$\min_{x,y} \quad G_1(x) + G_2(y), \quad \text{subject to} \quad Ax - y = 0, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m$$

Alternating direction method of multipliers

- Decomposition algorithm for solving convex optimization problems
- Based on iterative scatter and gather operations on the augmented Lagrangian
- Solves problems of the form:

$$\min_{x,y} \quad G_1(x) + G_2(y), \quad \text{subject to} \quad Ax - y = 0, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m$$

- Augmented Lagrangian:

$$L_\rho(x, y, z) = G_1(x) + G_2(y) + z^T(Ax - y) + \rho/2 \, \|Ax - y\|_2^2$$

Alternating direction method of multipliers

- Decomposition algorithm for solving convex optimization problems
- Based on iterative scatter and gather operations on the augmented Lagrangian
- Solves problems of the form:

$$\min_{x,y} \quad G_1(x) + G_2(y), \quad \text{subject to} \quad Ax - y = 0, \quad x \in \mathbb{R}^n, \quad y \in \mathbb{R}^m$$

- Augmented Lagrangian:

$$L_\rho(x, y, z) = G_1(x) + G_2(y) + z^T(Ax - y) + \rho/2 \, \|Ax - y\|_2^2$$

- Iterative update equations:

$$\begin{aligned} x^{t+1} &= \min_x \left\{ G_1(x) + z^{tT} Ax + \rho/2 \left\| Ax - y^t \right\|_2^2 \right\} \\ y^{t+1} &= \min_y \left\{ G_2(y) - z^{tT} y + \rho/2 \left\| Ax^{t+1} - y \right\|_2^2 \right\} \\ z^{t+1} &= z^t + \rho \left(Ax^{t+1} - y^{t+1} \right) \end{aligned}$$

SPI-GP using ADMM

$$\min \quad \mathbf{Tr}(KS) - \log det(S) + \lambda \|Y\|_1 \quad \text{subject to} \quad S - Y = 0$$

SPI-GP using ADMM

$$\min \quad \mathbf{Tr}(KS) - \log det(S) + \lambda \left\| Y \right\|_1 \quad \text{subject to} \quad S - Y = 0$$

Iterative equations:

$$\begin{aligned}
S^{t+1} &= \min_x (\mathbf{Tr}(KS) - \log det(S) + \rho/2 \left\| S - Y^t + P^t \right\|_F) \\
Y^{t+1} &= \min_y \left(\lambda \left\| Y \right\|_1 + \rho/2 \left\| S^{t+1} - Y + P^t \right\|_F\right) \\
P^{t+1} &= P^t + \left(S^{t+1} - Y^{t+1}\right)
\end{aligned}$$

SPI-GP using ADMM

- Further simplification:

$$S^{t+1} = Q\widehat{S}Q^T$$

  - where $[Q \quad \Lambda] = eig(\rho(Y^t - P^t) - K)$, $\widehat{S}_{ii} = \frac{\lambda_i + \sqrt{\lambda_i^2 + 4\rho}}{2\rho}$

- Also,

$$Y_{ij}^{t+1} = SoftThreshold_{\lambda/\rho}\left(S_{ij}^{t+1} + P_{ij}^t\right)$$

Illustrative example

- Data: A 21 year (1982-2002) climate data consisting of NCEP/NCAR features cross-matched with data from NOAA/AVHRR
  - multiple features, monthly averages, recorded at 0.5° resolution over the earth's surface
- Example regression problem: Precipitation prediction in the Indian subcontinent in the month of August based on past precipitation data
  - Accuracy measure: Normalized mean squared error
- Results

	1982	1986	1990	1994	1998
Full-GP	0.695	0.664	0.611	0.651	0.669
SPI-GP	0.6912	0.664	0.605	0.650	0.667

Table: NMSE of Gaussian Process regression.

Experimental results: Precipitation network in SPI-GP

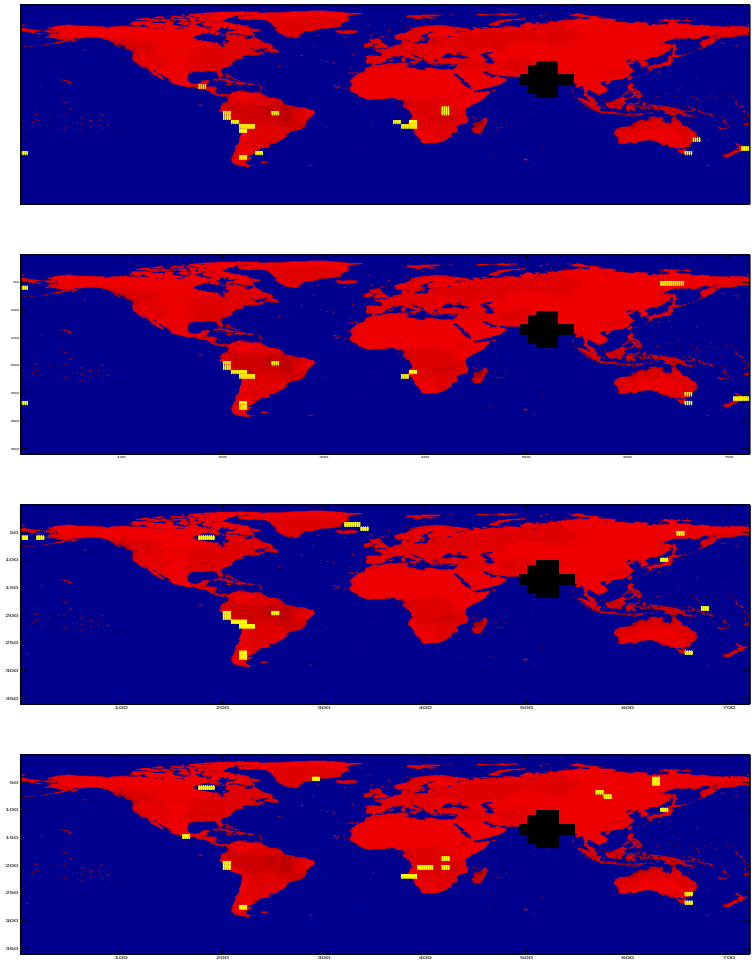


Figure: Evolution of the climate network over 20 years based on precipitation

Experimental results: scalability

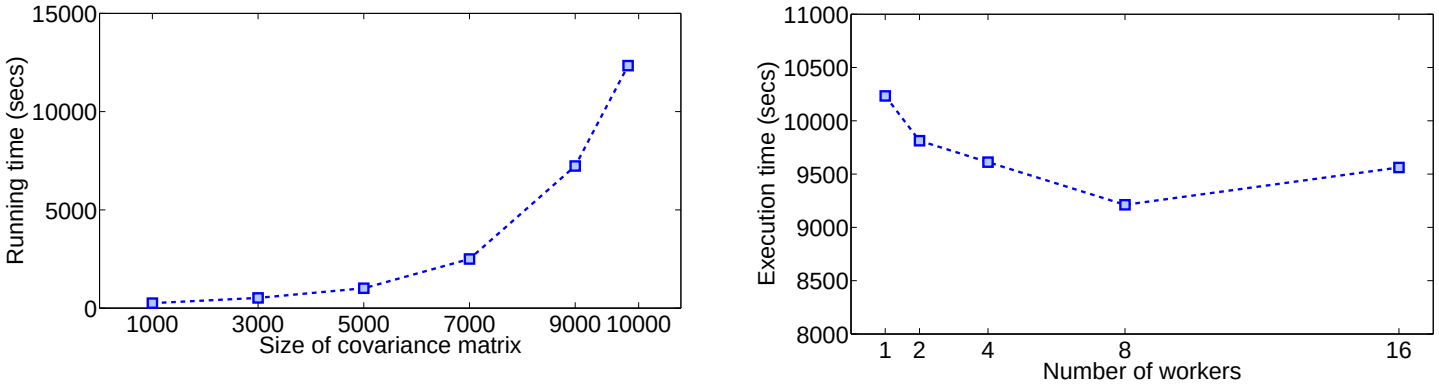


Figure: Scalability of SPI-GP using ADMM

Summary

- Developed a method for sparse Gaussian Process regression that allows us to build a parsimonious (interpretable) model for large data sets
- Replaced the kernel inversion operation with a distributable optimization based inverse estimation
- Demonstrated a good balance of accuracy and interpretability for the Gaussian Process regression models.
- Future work is to look at other inverse covariance estimation techniques and apply this method on specific climate domain applications in collaboration with domain scientists