`arleGP`: Approximated Restricted Likelihood Estimator for Large-Scale Spatial Gaussian Process in R

Zhou Lan, Rui Li, Chengchun Shi
North-Carolina State University
`cshi4@ncsu.edu`

Package: `arleGP`

- **Title**: Approximated Restricted Likelihood Estimator for Gaussian Process
- **Description**: This package provides parameter estimates of covariance structure for spatial Gaussian Process by maximizing the approximated restricted likelihood (Stein et al., 2004). The major optimization function is written in C with GNU Scientific Library (GSL, Galassi et al., 2015) to facilitate the computation.
- **System Requirement:** GNU Scientific Library version >= 1.8

Currently we only support Gaussian process with covariance structure

$$\mathrm{cov}(Z(x), Z(y)) = -\theta_2 \exp\left(-\frac{\theta_1 \|x-y\|_2}{\theta_2}\right).$$

Other covariance structures will be added later to the package.

Methods

Assume covariance matrix $K(\theta)$ is positive definite. Let Z_i be the ith observation, $Z_{(i)} = (Z_1, \ldots, Z_i)^T$, $W_i(\theta)$ the error of the BLUP of Z_i based on $Z_{(i-1)}$, $V_i(\theta)$ variance of $W_i(\theta)$.

Exact restricted log-likelihood (RL)

The restricted log-likelihood of θ in terms of Z is given by

$$\mathrm{rl}(\theta, Z) = -\frac{n-p}{2}\log(2\pi) - \frac{1}{2}\sum_{i=2}^{n}[\log\{\det(V_i)\} + W_i^2/V_i], \quad (1)$$

where p is the dimension of predictors, n number of observations.

Calculation of (1) requires $O(n^3)$ operations. The idea is to calculate BLUP of Z_i based on $S_{(i-1)}$ which is a subvector of $Z_{(i-1)}$, to simplify the calculation. Let $W_i(S)$ be error of BLUP of Z_i based on $S_{(i-1)}$, $V_i(S)$ its covariance matrix.

Approximated restricted log-likelihood

Instead of maximizing (1), we maximize

$$\mathrm{rl}(\theta, Z) = -\frac{n-p}{2}\log(2\pi) - \frac{1}{2}\sum_{i=2}^{n}[\log\{\det(V_i(S))\} + W_i(S)^2/V_i(S)]. \quad (2)$$

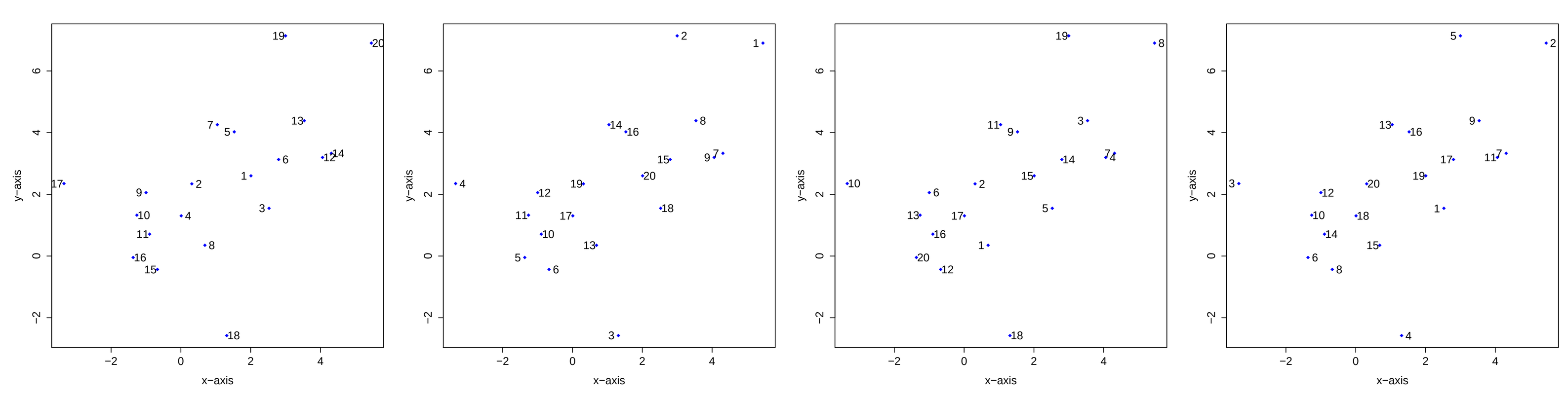
Maximizing (2) is equivalent to setting its derivative to be equal to 0. It's shown in Stein et al. (2004) that the expectation of the derivative is equal to 0. Well-developed theory of estimating equations can be applied to obtain standard errors of the proposed estimator.

Ordering

Ordering of the observations will affect the choice of conditioning subvectors, and hence has a non-negligble effect on the parameter estimates. `arleGP` provides five options to order the observations.

- "natural": order observations according to the sum of coordinates, i.e, from lower left-hand corner to the upper right-hand corner.
- "center-out", "outside-in": order observations according to their distances to the center
- "random": order observations randomly
- "max-min": assume we already pick k observations in a list. Choose the $k+1$th observation that maximizes the minimum distance among these k observations.

Figure 1: orders of observations, from left to right: "center-out", "outside-in", "random", "max-min"



Select the conditioning subset

Given the order of observations, for each Z_j, we choose the subset $S_{(j-1)}$ of length m that corresponds to the top m nearest neighbours of the jth observation. m is a user-specified argument.

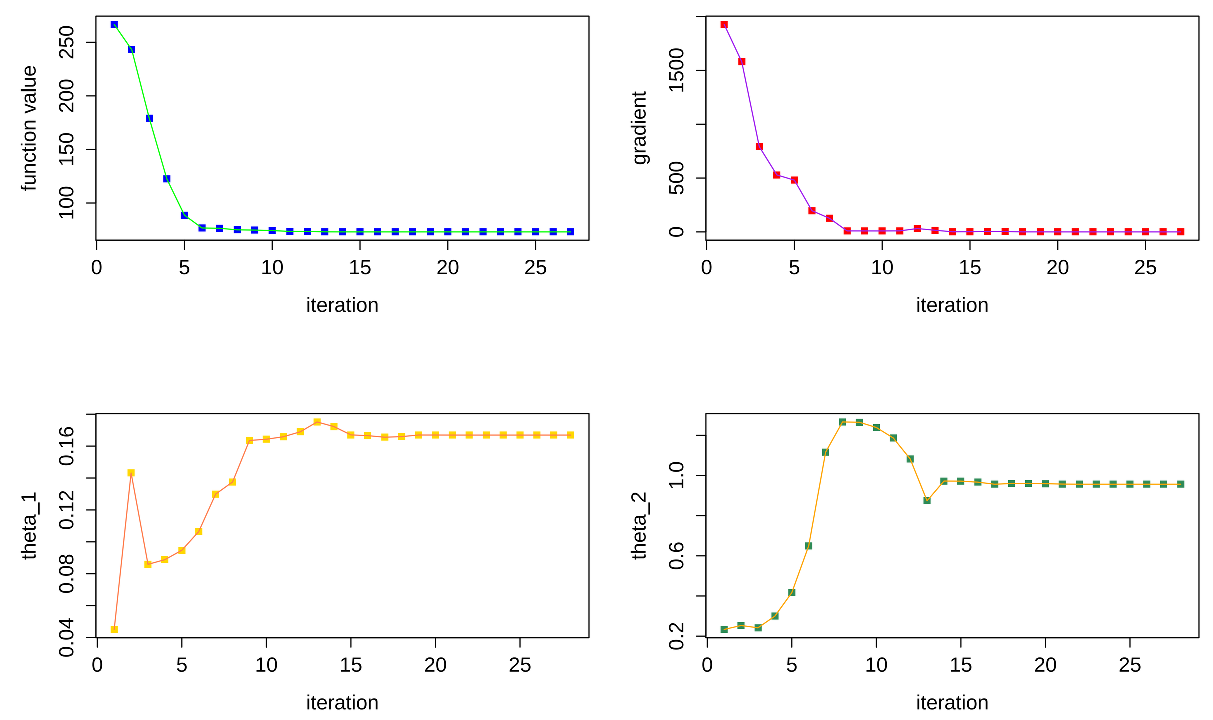
Functionality of `arleGP`

- **Exact RL inference**: `arleGP` by setting m (the length of conditioning vector) greater or equal to the sample size.
- **Approximated RL inference**: `arleGP`.
- **Specify mean structure**:
  1. `mean="zero"`: Gaussian process with zero mean
  2. `mean="constant"`: Gaussian process with constant mean
  3. `mean="linear"`: Gaussian process with mean modeled by linear regression

Fletcher-Reeves conjugate gradient algorithm is used when maximizing the (approximated) RL objective function.

Convergence of algorithm

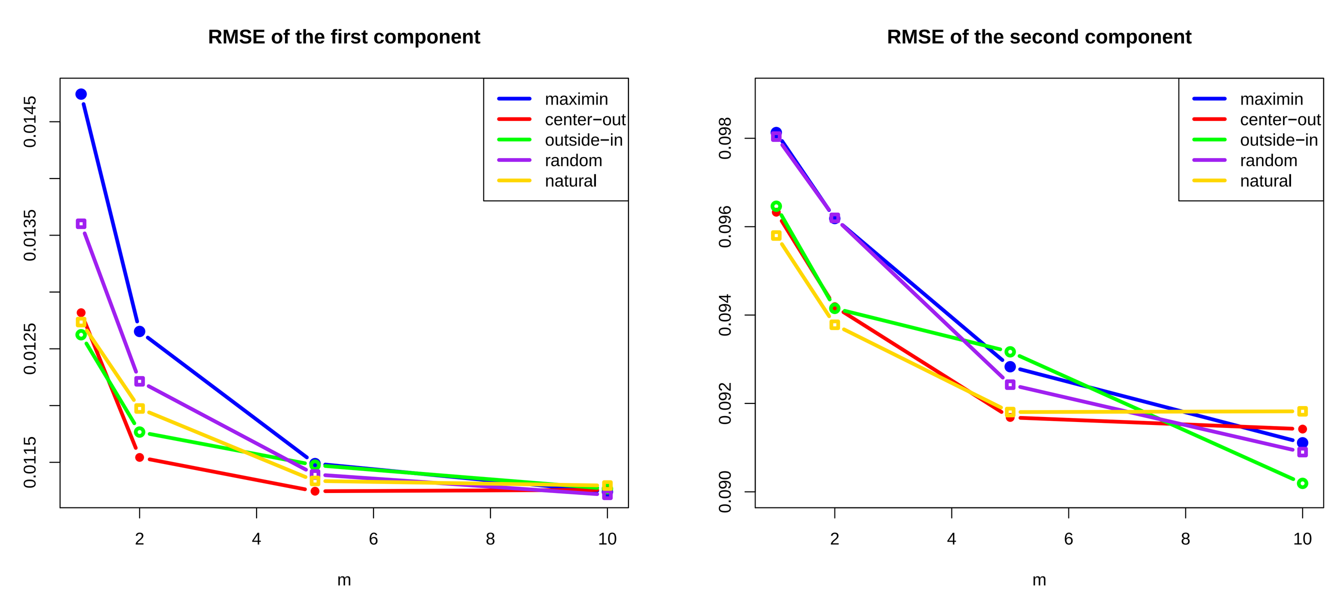
We use a toy example to examine convergence of our algorithm. Consider one observation network: 300 sites selected randomly out of 10000 points in the plane (i, j), $i, j = 1, \ldots, 100$, and then each perturbed by adding a random point in $[-0.25, 0.25]^2$. We set $\theta_1 = 0.2$, $\theta_2 = 1$ in the covariance structure.



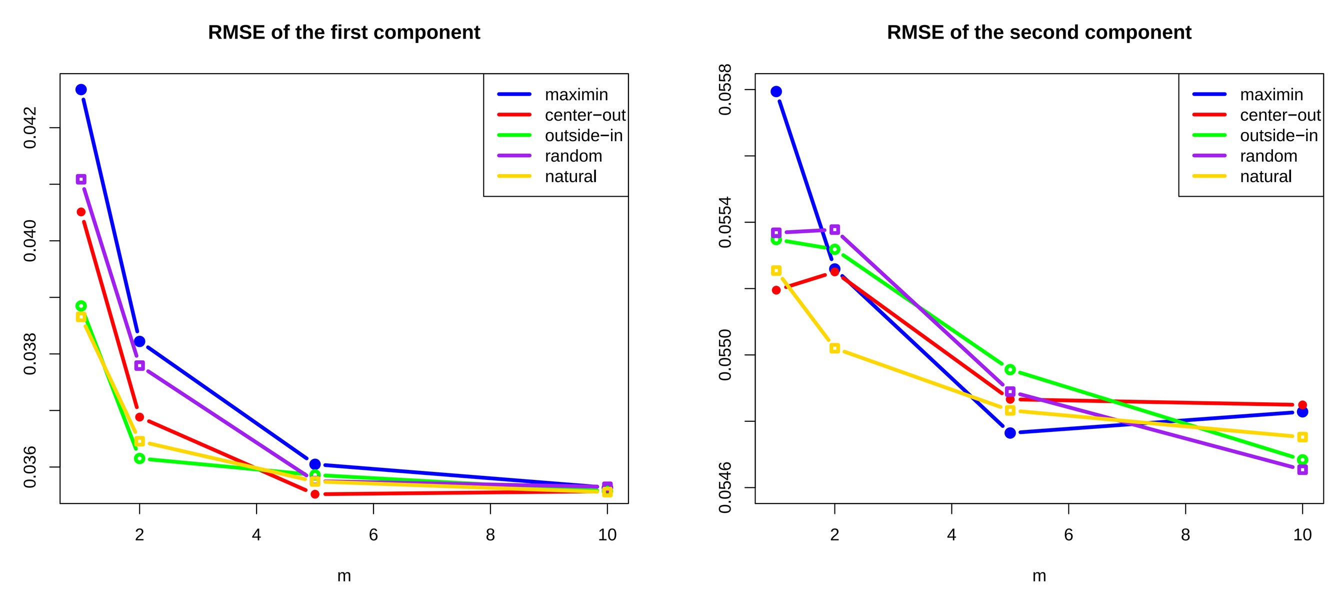
We choose $m = 30$. Function `arleGP` returns the objective value, gradient and current estimates at each iteration, which could be used to examine the convergence of the algorithm.

Simulation study

- Conduct simulation study to examine different ordering methods. Report root mean square error (RMSE) of parameter estimates.
- Scenario 1: $n = 1000$, $\theta_1 = 0.2$, $\theta_2 = 1$:



- Scenario 2: $n = 1000$, $\theta_1 = 0.5$, $\theta_2 = 1$:



- Conclusions:
  1. The orders do make a difference when m is small, but the difference is not significant.
  2. Estimators are very efficient even for small m.