

Instructions

Single-Sample simulation

In this repository, we have a set of functions to simulate data from a single matrix-variate normal distribution and run our proposed Blocked Gibbs Sampling algorithm. The R script **JOBS_Simulation.R** contains all necessary functions for the simulation. Sourcing the function loads these functions for use in the R environment. We use the function **generate_data** to simulate matrix-variate spatial data. The arguments for the function include

- `n_locs`: number of spatial locations.
- `d`: dimension of spatial locations. Defaults to 2. Usually 2 or 3.
- `range`: range of the Matern kernel used to simulate the true column covariance matrix.
- `smoothness`: smoothness parameter of the Matern kernel used to simulate the true column covariance matrix.
- `scale`: marginal variance of the Matern kernel used to simulate the true column covariance matrix.
- `r`: value of rho to simulate the scale matrix corresponding to the row covariance matrix. type: one of “AR”, “Band”, “Equi”. Denotes the type of row scale structure. Either AR correlation, Banded correlation or Equi correlation structure used to simulate the scale matrix corresponding to the row covariance matrix.

The main function to run the proposed Blocked Gibbs sampler is **posterior_est**. The arguments of the function are

- `data`: a list consisting of the spatial data, spatial locations, true log-likelihood etc. The data is from the `generate_data` function
- `m`: the number of neighbours
- `n_iterations`: number of MCMC iterations.
- `burn`: number of samples to be considered as burn-in
- `thin`: thinning factor of the MCMC samples

Note that this function, automatically stores the traceplot corresponding to the posterior log-likelihood for our model and the corresponding ACF plot. We may use these plots as diagnostic tools to understand the convergence properties of the proposed sampler. The traceplot also includes the true value of the log-likelihood corresponding to the simulated data as a reference.

Example

Data generation

For the purpose of illustration, we first simulate data from a matrix-variate normal distribution with 200 randomly generated spatial locations from the distribution, $Uni(0, 1)$ and 20 rows (corresponding to genes in spatial transcriptomic data). The true spatial covariance matrix is generated from a Matern kernel with smoothness 0.5, range 1.5 and marginal variance 1. The true scale matrix Ψ is generated from an AR structure i.e., with $\Psi_{i,j} = \rho^{|i-j|}$ where $\rho = 0.5$.

```
# First source the function JOBS_Simulation.R file, which loads all the relevant  
# R functions (both custom functions and inbuilt functions)  
suppressPackageStartupMessages(source("JOBS_Simulation.R"))
```

```
## Warning: package 'spam' was built under R version 4.3.1
```

```
## Warning: package 'Matrix' was built under R version 4.3.1
```

```
## Warning: package 'ggplot2' was built under R version 4.3.1
```

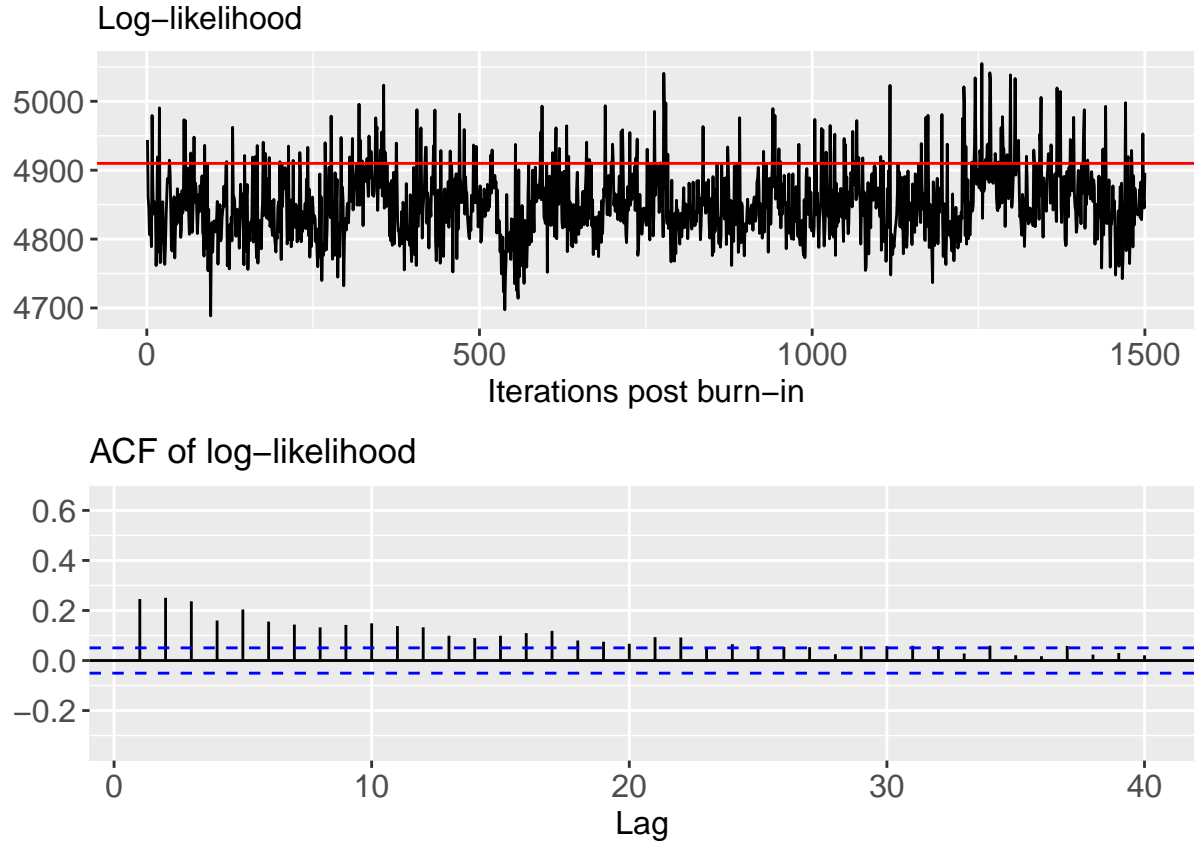
```
## Warning: package 'lubridate' was built under R version 4.3.1
```

```
datum <- generate_data(n_locs = 200, num_reps = 20, range = 1.5)
```

Running the proposed MCMC algorithm

We run our proposed Blocked Gibbs sampler for 2000 iterations and consider first 500 samples as burn-in. We do not consider any thinning of the samples. We plot the traceplot of log-likelihood for each of the posterior samples from our sampler and the corresponding auto-correlation plot. We overlay the traceplot of log-likelihood with the true value of the log-likelihood.

```
Blocked.MCMC = posterior_est(data = datum, n_iterations = 2000, burn = 500, thin = 1)
```



Multi-Sample simulation

We also provide the code for performing simulations for the multi-sample case. The associated R code is **JOBS_Replicated_Simulation.R**. For simplicity, we are not reproducing the simulations in the this instructions. The code may be run as it is for various sizes of spatial locations and number of correlated observations.

Real Data Analysis

The STARmap data used for the real data analysis can be found in the subfolder **STARmap**. The folder **Real Data Analysis** has the code **JOBS_STARmap.R** which consists of all the analysis for the STARmap data using our proposed method JOBS. This includes data pre-processing, running our Blocked-Gibbs sampler, and to produce all plots presented in the paper. The folder also contains the codes **DR.SC_STARmap.R** and **BayesSpace_STARmap.R** which looks at the analysis of STARmap data using two competing methods, namely, DR.SC and BayesSpace. These codes also produce the necessary plots as presented in the paper.