Plan of software development for smoothing and estimating functional data with **B**ayesian hierarchical Gaussian Process regression model with **a**pproximations by **b**asis **f**unctions (BABF)

1. Write function to generate simulation data from a , where covariance is a matern function with signal variance , scale parameter , and order that control smoothness of signals. Noise follows i.i.d. . Should allow given choices of all parameter values and mean function (if possible). Matlab code of Matern function is given as Matern.m; matlab code of generating signals on common grid is given as gaussian\_sim.m; and matlab code of generating signals on random grids is given as gaussian\_rsim.m.
2. Try first with common grid, equally spaced 40 grid points on . Then random grid, random 40 grid points uniformly generated from .
3. Always generate random multivariate-normal samples, by generating i.i.d. normal(0, 1) sample first, then add the mean, and multiply the Cholesky factor of the covariance matrix. Also generate inverse Wishart samples by generating multivariate-normal samples first, and then do the transformation by yourself, if you do not find an existing function to use.
4. Please use existing C++ scientific libraries (e.g., GSL, EIGEN) for matrix storing and calculations.
5. Generate simulation data from a non-stationary covariance function, following matlab code of gaussian\_simNS.m and gaussian\_rsim\_ns.m.
6. Try to write a function to obtain **smooth estimates of the mean and covariance matrix** of the functional signals (start with common-grid case). You will need both estimates for the BABF algorithm.
7. Write functions to convert a non-positive-definite covariance matrix to positive-definite, following topdm.m.
8. Try to write a function or call a C++ library function to obtain **cubic-spline basis functions for given knots**. You will need cubic-spline basis functions for the BABF algorithm.
9. Try to code the BABF method, following matlab code spmcmc.m.
10. Output should be similar to the matlab code spmcmc.m, mainly **estimates along with 95% CIs for all parameters in the model**. Should allow choices of outputs.
11. Output should be written as txt files, and the software should allow inputs of txt files about functional signals, mean and covariance estimates. Of course, allow inputs of other parameter values, as well as options of outputs.

Comments:

* Start with generating simulation data with Matern covariance function (stationary case) and common-grid.
* You should look at what are the best data formats for functional signals and the outputs.
* You should think about how to organize your code.
* You should use well-known numerical C++ libraries when possible, e.g., GSL (http://www.gnu.org/software/gsl/), EIGEN (http://eigen.tuxfamily.org/index.php?title=Main\_Page)…. Rmath is a standalone library, which allows you to call all R functions. But you need to figure out how to generate the standalone library and call it in your code (I put an introduction in the Reference folder).
* Github is only a place to store your source code. You could do all the development and testing on your local PC, and push your final source code to Github when everything is OK. If you need to test something, you can always start a new branch from the main branch using git; then push good resulting code back to main branch.
* For more details about the algorithm and simulation setups, please look at the manuscript draft (under folder manuscript in the shared dropbox folder).
* All matlab codes are located under folder BFA\_Code in the shared dropbox folder.
* Please take a look at the files I put under SmoothingCpp/References folder about how to call R functions in C++.
* README.md is like a manual for this software.
* You could put example data under SmoothingCpp/example folder, and put any testing outputs under SmoothingCpp/out folder.