B-MASTER Reproducibility Instructions

We describe the reproducibility instructions in three sections. Please refer to the following list to see which section describes the reproducibility of the tables and figures below.

Main Figures (12 items)

- Figure 1: Concept diagram, no coding involved.
- Figure 2: refer to **Real data analysis**.
- Figure 3: Concept diagram, no coding involved.
- Figure 4: refer to **Simulation study 2**.
- Figure 5: refer to **Simulation study 1**.
- Figure 6: refer to **Simulation study 2**.
- Figure 7 - 12: refer to **Real data analysis**.

Supplementary Figures (6 items)

- Figure S1: Same as main Figure 5.
- Figure S2: Same as main Figure 6.
- Figure S3: refer to **Simulation study 1**.
- Figure S4-S6: refer to **Real data analysis**.

Main Tables (2 items)

- Table 1: refer to **Simulation study 1**.
- Table 2: refer to **Real data analysis**.

Supplementary Tables (7 items)

- Table S1: Notation table, no coding involved.
- Table S2: Same as main Table 1.
- Table S3: refer to **Simulation study 2**.
- Table S4: refer to **Simulation study 1**.
- Table S5: refer to **Simulation study 1**.
- Table S6, S7: refer to **Real data analysis**.

1. Simulation study 1

First we go through the main simulation study steps.

1. **Create Dummy data:** Go to Simulation study 1 / Dummy Real Data / Generate_Dummy_Real_via_BMASTER.
2. Run BMASTER_MockRealDataGenerate.m to generate estimated coefficient matrix, which is considered as the True coefficient matrix for the rest of this simulation study.
3. Outputs are generated within Generate_Dummy_Real_via_BMASTER / Data.
4. The generated data is copied and placed in Simulation study 1 / Dummy Real Data (already copied, no need to copy again).
5. In Dummy Real Data, run GENERATE_Y_DATA.m to generate 10 realizations of Y based on X and earlier estiamted B ("True beta"). Now data generation is completed.
6. Copy the generated datasets to the sub-folders within Simulation study 1, namely, BMASTER, SSLasso, mSSL, Remmap (already copied, no need to copy again).
7. **B-MASTER results:** Run BMASTER / BMASTER_on_simReal_v3.m.
8. **SSLasso results:** Run SSLasso / SSLASSO_on_simReal.R.
9. **mSSL results:** Run mSSL / mSSL_dpe_on_simReal.R (for `dpe` method), and mSSL / mSSL_dCpe_on_simReal.R (for `dcpe` method).
10. **Remmap results:** Run Remmap / remMAP_on_simReal.R (for original method), and Remmap / remMAPBic_on_simReal.R (for BIC-based approximation method).
11. Copy all the outputs to Simulation study 1 / Summary Table and Plots (already copied, no need to copy again).

12. Run Summary Table and Plots / Summary_tables_plots.R to generate **Table 1** outputs, and **Figure 5 (bottom right)**.

13. Run Summary Table and Plots / Post_analysis_plot.R to generate **Figure 5 (up; bottom left)**.

Now we detail the posterior convergence and hyperparameter sensitivity diagnosis.

1. Go to Simulation study 1 / BMASTER / Sensitivity and convergence analysis.

2. Run BMASTER_on_simReal_convergence_check.m, which generates **Figure S3** and the diagnostic measures noted in **Table S4**. The generated figure is saved within · / Sensitivity and convergence analysis / Output folder.

3. Run BMASTER_on_simReal_sensitivity.m once for each scenarios `(delta_1,delta_2)` = (0.01,0.01), (0.01,1), (1, 0.01), (1,1). Corresponding raw outputs are saved within · / Sensitivity and convergence analysis / Output folder.

4. Run Summary_sensitivity.m, which processes the raw outputs generated from previous code, summarizes and saves final output files in · / Sensitivity and convergence analysis / Output. That is presented in **Table S5**.

2. Simulation study 2

1. Go to Simulation study 2 and open BMASTER_scalability.m.

2. Setting `IsRhoNonZero` $= 0$ run it for P= $(20, 50, 100, 200, 500, 1000, 2000)$.

3. Setting `IsRhoNonZero` $= 1$ run it for P= $(20, 50, 100, 200, 500, 1000, 2000)$.

4. Run Simulation study 2 / Summary_comp_time.R, that generates **Figure 6**, and two csv files corresponding to the upper and lower halves of **Table S3**.

5. Go to Simulation study 2 / remMap scalability.

6. Run BMASTER_v_remmap.m thrice setting `SampleMultFactor` $= 1, 5, 10$.

7. Run remMAP_scalability.R thrice setting `SampleMultFactor` $= 1, 5, 10$.

8. Run Comparison_plot.R. That generates **Figure 4**.

3. Real data analysis

First, we describe the generation process for the exploratory data analysis plots and tables.

1. Go to Real Data Analysis / Data / Yachida_BMASTER. Run 1_Full_raw_data_extraction_only.R for first level data extraction, then run 2_Data_analysis.R to perform second level data extraction; this also yields **Figure 2(c)**. Some data processing related details from the output of aforementioned codes are summarized in the concept diagram **Figure 2(a)**.

2. Run 3_Extra_plots.R to reproduce **Figure 2(b), (d), (e), (f)** (that completes all subplots under **Figure 2**).

3. Run 4_Extra_supp_plots.R to reproduce **Figure S5, S6**.

4. Run 5_patient_summary_plots.R to reproduce **Figure S4**.

5. Run 6_Scores_plot.R to yield **Figure 7**.

Now we note down steps to reproduce other tables and plots generated after fitting B-MASTER.

1. Go to Real Data Analysis. Run BMASTER_Real_Data.m.

2. Run Extraction_subset_for_analysis.R.

3. Run Overall_FIS_plots.R. That generates **Figure 8, 9**.

4. Run Post_analysis_subset_1_NEW.R. That generates **Figure 10** and the Bayesian p-values noted in **Table S6**.

5. Run Post_analysis_subset_2_NEW.R. That generates **Figure 11** and the Bayesian p-values noted in **Table S7**.

6. Run CCA_subset_plots.R. That generates **Figure 12**.

7. Go to Real Data Analysis / Validation on real data.

8. Run BMASTER_Real_Data_validation.m. Also run SSLASSO_RealData_validation.R, mSSL_dCpe_RealData_validation.R, remMAPBic_RealData_validation.R. This sequence of executions generate 4 outputs csv files corresponding to corresponding methods; which results are assembled and presented in **Table 2**.