Appendix to "Local Gaussian Process Model for Large-scale Dynamic Computer Experiments" published in the Journal of Computational and Graphical Statistics

Ru Zhang, C. Devon Lin
Department of Mathematics and Statistics,
Queen's University, ON, Canada
and
Pritam Ranjan
Operations Management & Quantitative Techniques,
Indian Institute of Management Indore, MP, India

February 22, 2018

This document presents proof of Proposition 1 (of the main article) in Section A. Section B contains two algorithms (in the formal algorithm format) for fitting local approximate SVD-based GP models (lasvdGP) described in Sections 2 and 3 of the article. Section C summarizes simulation results for establishing the reliability of the estimated range parameters (or equivalently, the correlation parameters) for the proposed lasvdGP model fits in Examples 1 and 2 of the main article.

A. Proof Of Proposition 1

Following (7), the inner expectation in (10) can be written as

$$\begin{aligned}
&\mathrm{E}\Big[\big\|\boldsymbol{y}(\boldsymbol{x}_0)-\hat{\boldsymbol{y}}(\boldsymbol{x}_0|\boldsymbol{c}(\boldsymbol{x}),\boldsymbol{V}^{*(k)},\hat{\boldsymbol{\Theta}}^{(k)})\big\|^2\Big|\boldsymbol{c}(\boldsymbol{x}),\boldsymbol{V}^{*(k)},\hat{\boldsymbol{\Theta}}^{(k)},(\hat{\sigma}^{(k)})^2\Big]\\
=&\mathrm{tr}\Big(\boldsymbol{B}^{(k)}\boldsymbol{\Lambda}\big(\boldsymbol{V}^{*(k)}(\boldsymbol{x}),\hat{\boldsymbol{\Theta}}^{(k)}\big)(\boldsymbol{B}^{(k)})^T+(\hat{\sigma}^{(k)})^2\boldsymbol{I}_L\Big)\\
=&(\hat{\sigma}^{(k)})^2L+\mathrm{tr}\Big(\boldsymbol{\Lambda}\big(\boldsymbol{V}^{*(k)}(\boldsymbol{x}),\hat{\boldsymbol{\Theta}}^{(k)}\big)(\boldsymbol{B}^{(k)})^T\boldsymbol{B}^{(k)}\Big)\\
=&(\hat{\sigma}^{(k)})^2L+\sum_{i=1}^{p_k}(d_i^{(k)})^2\hat{\sigma}_i^2\big(\boldsymbol{x}_0|\boldsymbol{v}_i^{(k)}(\boldsymbol{x}),\hat{\boldsymbol{\theta}}_i^{(k)}\big),
\end{aligned}\tag{A.1}$$

where $\boldsymbol{V}^{*(k)}(\boldsymbol{x})=[(\boldsymbol{V}^{*(k)})^T,\boldsymbol{c}(\boldsymbol{x})]^T$ and $d_i^{(k)}$ is the ith largest singular value of $\boldsymbol{Y}^{(k)}$,

$$\boldsymbol{\Lambda}\big(\boldsymbol{V}^{*(k)}(\boldsymbol{x}),\hat{\boldsymbol{\Theta}}^{(k)}\big)=\mathrm{diag}\Big(\hat{\sigma}_1^2\big(\boldsymbol{x}_0|\boldsymbol{v}_1^{(k)}(\boldsymbol{x}),\hat{\boldsymbol{\theta}}_1^{(k)}\big),\ldots,\hat{\sigma}_{p_k}^2\big(\boldsymbol{x}_0|\boldsymbol{v}_{p_k}^{(k)}(\boldsymbol{x}),\hat{\boldsymbol{\theta}}_{p_k}^{(k)}\big)\Big),$$

and

$$\hat{\sigma}_i^2\big(\boldsymbol{x}_0|\boldsymbol{v}_i^{(k)}(\boldsymbol{x}),\hat{\boldsymbol{\theta}}_i^{(k)}\big)=\frac{\rho_i^{(k)}(\boldsymbol{x}_0,\boldsymbol{x})}{\alpha_i+k}\big(\beta_i+\psi_i^{(k)}(\boldsymbol{x})\big),$$

for $i=1,\ldots,p_k$, where $\psi_i^{(k)}(\boldsymbol{x})=\boldsymbol{v}_i^{(k)}(\boldsymbol{x})^T\tilde{\boldsymbol{K}}_i^{-1}(\boldsymbol{x})\boldsymbol{v}_i^{(k)}(\boldsymbol{x})$, and $\boldsymbol{v}_i^{(k)}(\boldsymbol{x})=[(\boldsymbol{v}_i^{(k)})^T,c_i(\boldsymbol{x})]^T$ is the ith column of $\boldsymbol{V}^{*(k)}(\boldsymbol{x})$.

The first equality of (A.1) follows from Theorem 3.2b.1 of Mathai and Provost (1992). The third equality is derived from the column-orthogonality of $\boldsymbol{B}^{(k)}$, i.e. $(\boldsymbol{B}^{(k)})^T\boldsymbol{B}^{(k)}=(\boldsymbol{D}^{*(k)})^2$. Plugging (A.1) into (10), we get

$$\begin{aligned}
J(\boldsymbol{x}_0,\boldsymbol{x})&=\mathrm{E}\Big[(\hat{\sigma}^{(k)})^2L+\sum_{i=1}^{p_k}(d_i^{(k)})^2\hat{\sigma}_i^2\big(\boldsymbol{x}_0|\boldsymbol{v}_i^{(k)}(\boldsymbol{x}),\hat{\boldsymbol{\theta}}_i^{(k)}\big)\Big|\boldsymbol{V}^{*(k)},\hat{\boldsymbol{\Theta}}^{(k)},(\hat{\sigma}^{(k)})^2\Big]\\
&=(\hat{\sigma}^{(k)})^2L+\sum_{i=1}^{p_k}(d_i^{(k)})^2\Big(\frac{\rho_i^{(k)}(\boldsymbol{x}_0,\boldsymbol{x})}{\alpha_i+k}\big(\beta_i+\mathrm{E}[\psi_i^{(k)}(\boldsymbol{x})|\boldsymbol{V}^{*(k)},\hat{\boldsymbol{\Theta}}^{(k)},(\hat{\sigma}^{(k)})^2]\big)\Big)\\
&=(\hat{\sigma}^{(k)})^2L+\sum_{i=1}^{p_k}(d_i^{(k)})^2\Big(\frac{\rho_i^{(k)}(\boldsymbol{x}_0,\boldsymbol{x})}{\alpha_i+k}\big(\beta_i+\mathrm{E}[\psi_i^{(k)}(\boldsymbol{x})|\boldsymbol{v}_i^{(k)},\hat{\boldsymbol{\theta}}_i^{(k)}]\big)\Big)\\
&=(\hat{\sigma}^{(k)})^2L+\sum_{i=1}^{p_k}(d_i^{(k)})^2\Big(\frac{\rho_i^{(k)}(\boldsymbol{x}_0,\boldsymbol{x})}{\alpha_i+k}\big(\beta_i+\frac{\alpha_i+k}{\alpha_i+k-1}\psi_i^{(k)}\big)\Big).
\end{aligned}$$

The second equality holds because $\rho_i^{(k)}(\boldsymbol{x}_0,\boldsymbol{x})$ is a deterministic function of $\boldsymbol{x}_0$, $\boldsymbol{x}$ and $\hat{\boldsymbol{\theta}}_i^{(k)}$. The third equality follows from the independence among c_i's. The validity of the fourth equality is due to Gramacy and Apley (2015).

B. Algorithms

Algorithm 1 summarizes the key steps required for estimating the necessary parameters in the posterior predictive distribution (Equation (9) of the main article) of a full SVD-based GP model fitted to a training data of size N.

Algorithm 1: SVD-based GP model

Input : (1) Training set: $\boldsymbol{X}_{N\times q}$, (2) response matrix: $\boldsymbol{Y}_{L\times N}$, (3) threshold γ, (4) prior parameters: $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p, \alpha]^T$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p, \beta]^T$.

Output: (1) Basis $\boldsymbol{B}_{N\times p}$, (2) singular values $\boldsymbol{D}^*_{p\times p}$, (3) coefficients $\boldsymbol{V}^*$, (4) correlation parameters $\hat{\boldsymbol{\Theta}}$, (5) variance $\hat{\sigma}^2$.

Function `svdGP(`$\boldsymbol{X}$,$\boldsymbol{Y}$,$\boldsymbol{\alpha}$,$\boldsymbol{\beta}$,γ`)`

- $[\boldsymbol{B}, \boldsymbol{D}^*, \boldsymbol{V}^*, p] \leftarrow$ `buildBasis(`$\boldsymbol{Y}$,γ`)`
- $\boldsymbol{r} \leftarrow \text{vec}(\boldsymbol{Y}) - (\boldsymbol{I}_N \otimes \boldsymbol{B})\text{vec}(\boldsymbol{V}^{*T})$
- $\hat{\sigma}_i^2(\boldsymbol{x}_0|\boldsymbol{v}_i, \boldsymbol{\theta}_i) = (\beta_i + \psi_i)\Big(1 - \boldsymbol{k}_i^T(\boldsymbol{x}_0)\boldsymbol{K}_i^{-1}\boldsymbol{k}_i(\boldsymbol{x}_0)\Big)/(\alpha_i + N)$, $\hat{\sigma}^2 \leftarrow \big(\boldsymbol{r}^T\boldsymbol{r} + \beta\big)/(NL + \alpha + 2)$
- $\hat{\boldsymbol{\Theta}} \leftarrow$ `inference(`$\boldsymbol{V}^*$, p, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$`)`
- **return** $\boldsymbol{B}$, $\boldsymbol{D}^*$, $\boldsymbol{V}^*$, $\hat{\boldsymbol{\Theta}}$, $\hat{\sigma}^2$

Subroutine `buildBasis(`$\boldsymbol{Y}$,γ`)`

- $[\boldsymbol{U}, \boldsymbol{D}, \boldsymbol{V}] \leftarrow$`SVD(`$\boldsymbol{Y}$`)` `/* perform SVD on matrix` $\boldsymbol{Y}$`. */`
- $p \leftarrow \min\left\{m : \frac{\sum_{i=1}^m d_i}{\sum_{i=1}^k d_i} > \gamma\right\}$ `/* where` $\boldsymbol{D} = \text{diag}(d_1, \ldots, d_N), k = \min\{N, L\}$ `*/`
- $\boldsymbol{B} \leftarrow \boldsymbol{U}^*\boldsymbol{D}^*$ `/* as in Section 2.1 */`
- **return** $\boldsymbol{B}$, $\boldsymbol{D}^*$, $\boldsymbol{V}^*$, p

Subroutine `inference(`$\boldsymbol{V}^*$, p, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$`)`

- **for** $i \leftarrow 1$ **to** p **do**
  - $\hat{\boldsymbol{\theta}}_i \leftarrow \underset{\boldsymbol{\theta}_i}{\text{argmax}}\, \pi(\boldsymbol{\theta}_i|\boldsymbol{v}_i)$ `/* fit` p `independent GPs by finding the MAPs */`
- **return** $\hat{\boldsymbol{\Theta}} = [\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_p]^T$

Algorithm 2 presents the steps required for fitting the proposed local approximate SVD-based GP model (lasvdGP) with the neighbourhood points selected using the J-criterion in Section 3.1 of the main article.

Algorithm 2: Proposed local SVD-based GP model

Input : (1) Training set: $\boldsymbol{X}_{N\times q}$, (2) response matrix: $\boldsymbol{Y}_{L\times N}$, (3) test set $\boldsymbol{X}^*_{M\times q}$, (4) neighborhood size n, (5) initial neighborhood size n_0, (6) threshold γ, (7) prior parameters $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_p, \alpha]^T$ and $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_p, \beta]^T$.

Output: (1) The predicted mean response, and (2) the associated posterior variance in estimating $\boldsymbol{y}(\boldsymbol{x}_0)$ for each $\boldsymbol{x}_0 \in \boldsymbol{X}^*$.

---

for *each* $\boldsymbol{x}_0 \in \boldsymbol{X}^*$ **do**

  $\boldsymbol{X}^{(n_0)} \leftarrow \{\boldsymbol{x}_i, i = 1, \ldots, n_0\}$ `/* ` n_0 `nearest neighbours of` $\boldsymbol{x}_0$ `in` $\boldsymbol{X}$ `as in knn */`

  $\boldsymbol{Y}^{(n_0)} \leftarrow \{y(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}^{(n_0)}\}$

  for $k \leftarrow n_0$ **to** $n-1$ **do**

    $[\boldsymbol{B}^{(k)}, \boldsymbol{D}^{*(k)}, \boldsymbol{V}^{*(k)}, p_k, \hat{\boldsymbol{\Theta}}^k, (\hat{\sigma}^{(k)})^2, (\hat{\boldsymbol{\sigma}}^{(k)})^2] \leftarrow$`svdGP(`$\boldsymbol{X}^{(k)}$`,`$\boldsymbol{Y}^{(k)}$`,`$\boldsymbol{\alpha}$`,`$\boldsymbol{\beta}$`,`$\gamma$`)`

    $\boldsymbol{x}^*_{k+1} \leftarrow \underset{\boldsymbol{x}\in\boldsymbol{X}\setminus\boldsymbol{X}^{(k)}}{\operatorname{argmin}} J(\boldsymbol{x}_0, \boldsymbol{x})$

    $\boldsymbol{X}^{(k+1)} \leftarrow \boldsymbol{X}^{(k)} \cup \boldsymbol{x}^*_{k+1}$

    $\boldsymbol{Y}^{(k+1)} \leftarrow \boldsymbol{Y}^{(k)} \cup \boldsymbol{y}(\boldsymbol{x}^*_{k+1})$

  $[\boldsymbol{B}^{(n)}, \boldsymbol{D}^{*(n)}, \boldsymbol{V}^{*(n)}, p_n, \hat{\boldsymbol{\Theta}}^{(n)}, (\hat{\sigma}^{(n)})^2, (\hat{\boldsymbol{\sigma}}^{(n)})^2] \leftarrow$ `svdGP(`$\boldsymbol{X}^{(n)}$`,`$\boldsymbol{Y}^{(n)}$`,`$\boldsymbol{\alpha}$`,`$\boldsymbol{\beta}$`,`γ`)`

  Predict $\boldsymbol{y}(\boldsymbol{x}_0)$ through $\pi(\boldsymbol{y}(\boldsymbol{x}_0)|\boldsymbol{V}^{*(n)}, \hat{\boldsymbol{\Theta}}^{(n)}, (\hat{\sigma}^{(n)})^2, (\hat{\boldsymbol{\sigma}}^{(n)})^2)$ in Eqn. (9)

C. Additional Simulation Results

We now investigate the reliability of the estimated range parameters (or equivalently, the correlation parameters) for the proposed local approximate SVD-based GP model (lasvdGP) fits in Examples 1 (Forrester et al., 2008 – $q = 3, N = 10000, M = 2000$, $n = 40$ and $n_0 = 20$) and 2 (Bliznyuk et al., 2008 – $q = 5, N = 10000, M = 2000$, $n = 50$ and $n_0 = 25$) of the main article.

To explain the results, recall that for each point in the test set, the SVD-based GP model

fitted on the neighbourhood set is represented using a p-dimensional basis as in Equation (1), where p is selected by the cumulative percentage criterion (Equation (2)). That is, for each test point, p independent GP models for each $c_i(\mathbf{x})$ are fitted in the respective neighbourhood searched. The value of p may be different for different test points. The frequency table of the number of leading basis functions for 2,000 test points for each of the two examples are displayed in Table S.1.

	p						
	3	4	5	6	7	8	total
Example 1	266	1734	0	0	0	0	2000
Example 2	15	161	873	825	124	2	2000

Table S.1: Frequency of p among the 2,000 test points in Examples 1 and 2.

For simplicity, we only report the estimated range parameters in the GP models corresponding to $c_1(\boldsymbol{x})$, $c_2(\boldsymbol{x})$ and $c_3(\boldsymbol{x})$ from the final fits, i.e., after $n - n_0$ follow-up points were added. Figures S.1 and S.2 display the boxplots of those 2,000 estimates for each range parameter in Examples 1 and 2, respectively.

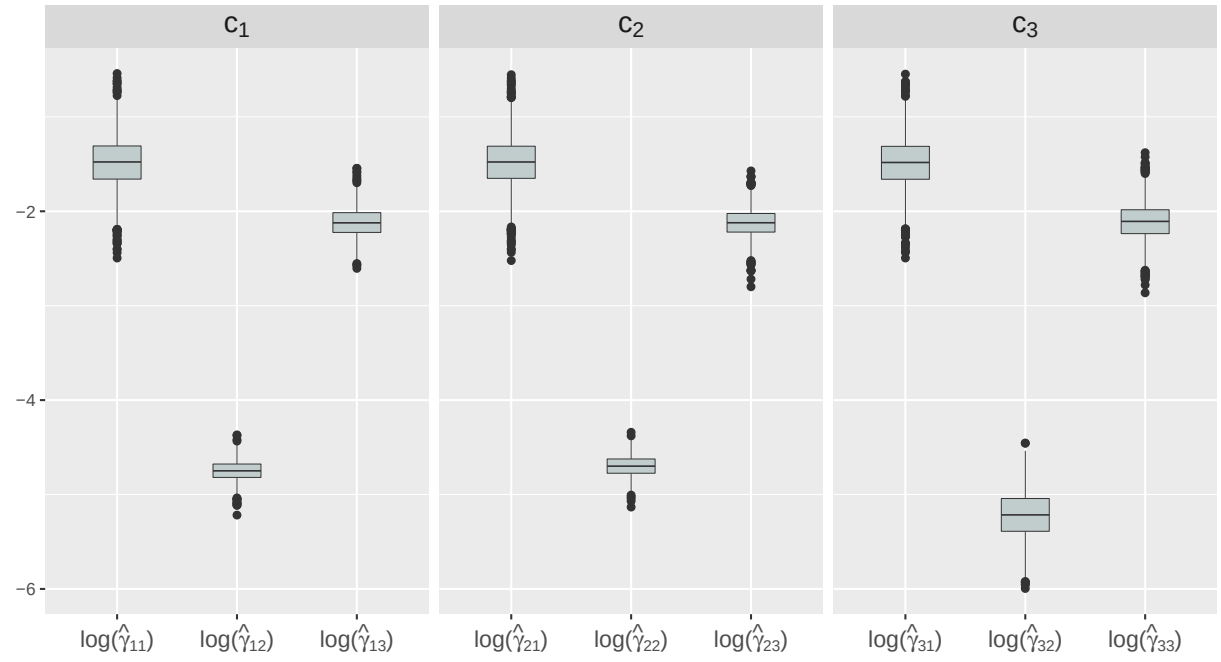

Figure S.1: The boxplots of the 2,000 estimates of the log-range parameters in the GP models for $c_1(\boldsymbol{x})$, $c_2(\boldsymbol{x})$ and $c_3(\boldsymbol{x})$ in Example 1 (Forrester et al., 2008).

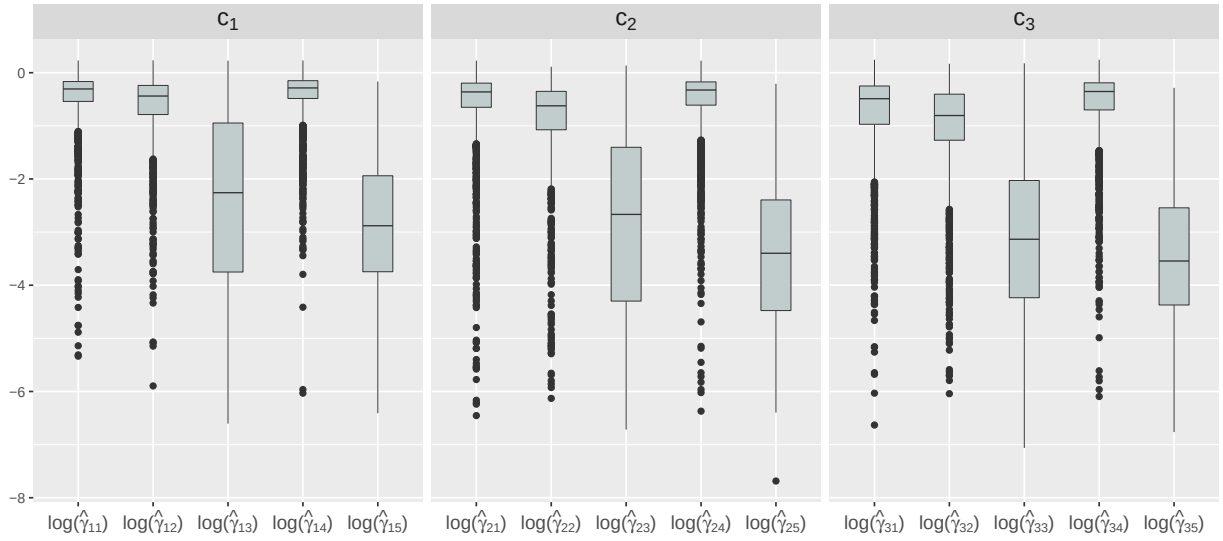

Figure S.2: The boxplots of the 2,000 estimates of the log-range parameters in the GP models for $c_1(\boldsymbol{x})$, $c_2(\boldsymbol{x})$ and $c_3(\boldsymbol{x})$ in Example 2 (Bliznyuk et al., 2008).