# Exploiting Variance Reduction Potential in Local Gaussian Process Search

Chih-Li Sung[a], Robert B. Gramacy[b], Benjamin Haaland[a]

[a]*Georgia Institute of Technology,* [b]*Virginia Tech*

## S1 Proof of Proposition 1

In the variance definition (4), the variance of $Y(x)$ at stage $j+1$ is

$$V_{j+1}(x) = \sigma^2\{\Phi_\Theta(x,x) - \Phi_\Theta(x, X_{j+1})\Phi_\Theta(X_{j+1}, X_{j+1})^{-1}\Phi(X_{j+1}, x)\}. \tag{S1.1}$$

Since $X_{j+1}$ is comprised of $X_j$ and $x_{j+1}$, (S1.1) can be rewritten as

$$V_{j+1}(x) = \sigma^2\Big\{\Phi_\Theta(x,x) - \begin{bmatrix}\Phi_\Theta(x,x_{j+1}) & \Phi_\Theta(x,X_j)\end{bmatrix}\begin{bmatrix}\Phi_\Theta(x_{j+1},x_{j+1}) & \Phi_\Theta(x_{j+1},X_j)\\ \Phi_\Theta(X_j,x_{j+1}) & \Phi_\Theta(X_j,X_j)\end{bmatrix}^{-1}\begin{bmatrix}\Phi_\Theta(x,x_{j+1})\\ \Phi_\Theta(X_j,x)\end{bmatrix}\Big\}. \tag{S1.2}$$

For simplicity, the second term of (S1.2) can be written as a partitioned matrix, that is,

$$\begin{bmatrix} a_1^T & a_2^T \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}^{-1} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{S1.3}$$

where

$$a_1 = \Phi_\Theta(x, x_{j+1}), a_2 = \Phi_\Theta(X_j, x),$$

$$B_{11} = \Phi_\Theta(x_{j+1}, x_{j+1}), B_{12} = \Phi_\Theta(x_{j+1}, X_j) = B_{12}^T \text{ and } B_{22} = \Phi_\Theta(X_j, X_j).$$

Applying partitioned matrix inverse results (Harville, 1997) and simplifying (S1.3) gives

$$a_2^T B_{22}^{-1} a_2 + (a_1 - B_{12} B_{22}^{-1} a_2)^T B_{11\cdot 2}^{-1} (a_1 - B_{12} B_{22}^{-1} a_2), \tag{S1.4}$$

where $B_{11\cdot 2} = B_{11} - B_{12} B_{22}^{-1} B_{21}$.

Then, taking (S1.4) into (S1.2) leads to

$$\begin{aligned} V(X_{j+1}) &= \sigma^2 \{\Phi_\Theta(x, x) - a_2^T B_{22}^{-1} a_2 - (a_1 - B_{12} B_{22}^{-1} a_2)^T B_{11\cdot 2}^{-1} (a_1 - B_{12} B_{22}^{-1} a_2)\} \\ &= \sigma^2 \{\Phi_\Theta(x, x) - \Phi_\Theta(x, X_j) \Phi_\Theta(X_j, X_j)^{-1} \Phi_\Theta(X_j, x) \\ &\quad - (a_1 - B_{12} B_{22}^{-1} a_2)^T B_{11\cdot 2}^{-1} (a_1 - B_{12} B_{22}^{-1} a_2)\} \\ &= V(X_j) - \sigma^2 \{(a_1 - B_{12} B_{22}^{-1} a_2)^T B_{11\cdot 2}^{-1} (a_1 - B_{12} B_{22}^{-1} a_2)\} \\ &= V(X_j) - \sigma^2 R(x_{j+1}), \end{aligned}$$

where

$$\begin{aligned} R(x_{j+1}) &= (a_1 - B_{12}B_{22}^{-1}a_2)^T B_{11\cdot 2}^{-1}(a_1 - B_{12}B_{22}^{-1}a_2) \\ &= (a_1 - B_{12}B_{22}^{-1}a_2)^2 / B_{11\cdot 2} \\ &= \frac{(\Phi_\Theta(x, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x))^2}{\Phi_\Theta(x_{j+1}, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x_{j+1})}, \end{aligned}$$

and the second equality holds since $B_{11\cdot 2}$ is a scalar.

## S2 Proof of Theorem 1

Since $(a-b)^2 \leq (a+b)^2$ for $a, b \geq 0$, equation (8) can be bounded as

$$\begin{aligned} R(x_{j+1}) &= \frac{(\Phi_\Theta(x, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x))^2}{\Phi_\Theta(x_{j+1}, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x_{j+1})} \\ &\leq \frac{(\Phi_\Theta(x, x_{n+1}) + \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x))^2}{\Phi_\Theta(x_{j+1}, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x_{j+1})}. \end{aligned}$$

Also, since

$$a^T B^{-1} b \leq \|a\|_2 \|B^{-1}b\|_2$$

and

$$a^T B^{-1} a \leq \|a\|_2^2 \lambda_{\max}(B^{-1}) = \|a\|_2^2 / \lambda_{\min}(B),$$

where $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ denote the maximum and minimum eigenvalues of a specific matrix, respectively, the inequality becomes

$$R(x_{j+1}) \leq \frac{(\Phi_\Theta(x, x_{j+1}) + \|\Phi_\Theta(x_{j+1}, X_j)\|_2 \|\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x)\|_2)^2}{1 - \|\Phi_\Theta(X_j, x_{j+1})\|_2^2 / \lambda_{\min}},$$

where $\lambda_{\min}$ is the minimum eigenvalue of $\Phi_\Theta(X_j, X_j)$.

Furthermore, according to the definition $d_{\min}(x_{j+1})$ of the minimum (Mahalanobis-like) distance as (6) and the definition $\phi(\cdot)$ as in Theorem 1, we have

$$\Phi_\Theta(u, x_{j+1}) \le \phi(d_{\min}(x_{j+1})), \text{ for any } u \in \{x, X_j\},$$

which also implies

$$\|\Phi_\Theta(x_{j+1}, X_j)\|_2 = \|\Phi_\Theta(X_j, x_{j+1})\|_2 \le \sqrt{j}\phi(d_{\min}(x_{j+1})),$$

therefore the inequality can be bounded as

$$R(x_{j+1}) \le \frac{(\phi(d_{\min}(x_{j+1})) + \sqrt{j}\phi(d_{\min}(x_{j+1}))\|\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x)\|_2)^2}{1 - j\phi^2(d_{\min}(x_{j+1}))/\lambda_{\min}}. \tag{S2.1}$$

Thus, for $\delta > 0$, if

$$\frac{(\phi(d_{\min}(x_{j+1})) + \sqrt{j}\phi(d_{\min}(x_{j+1}))\|\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x)\|_2)^2}{1 - j\phi^2(d_{\min}(x_{j+1}))/\lambda_{\min}} \le \delta$$

or equivalently

$$d_{\min}(x_{j+1}) \ge \phi^{-1}\left(\sqrt{\frac{\delta}{(1 + \sqrt{j}\|\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x)\|_2)^2 + j\delta/\lambda_{\min}}}\right),$$

then by (S2.1), $R(x_{j+1}) \le \delta$.

## S3 Proof of Theorem 2

Define $U(t) = (\sqrt{\lambda_1}\varphi_1(t), \sqrt{\lambda_2}\varphi_2(t), \ldots, \sqrt{\lambda_D}\varphi_D(t))^T \in \mathbb{R}^{D\times 1}$, where $\varphi_i(\cdot), i = 1, \ldots, D$ is an orthonormal basis of $L^2(\Omega)$ consisting of the eigenfunctions of $T$, defined in (13), and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_D$ are corresponding eigenvalues. According to (14), the approximated eigen-decomposition can be rewritten as

$$\Phi(x, y) \approx U^T(x)U(y).$$

Also, define a matrix $U(K) = [U(k_1), U(k_2), \ldots, U(k_n)] \in \mathbb{R}^{D\times n}$ for $K = (k_1, k_2, \ldots, k_n)$. Then, the reduction in variance $R(x_{j+1})$ in (8) can be approximated to the following:

$$\begin{aligned}
R(x_{j+1}) &= \frac{(\Phi_\Theta(x, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x))^2}{\Phi_\Theta(x_{j+1}, x_{j+1}) - \Phi_\Theta(x_{j+1}, X_j)\Phi_\Theta(X_j, X_j)^{-1}\Phi_\Theta(X_j, x_{j+1})} \\
&\approx \frac{(U^T(x_{j+1})U(x) - U^T(x_{j+1})U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)U(x))^2}{U^T(x_{j+1})U(x_{j+1}) - U^T(x_{j+1})U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)U(x_{j+1})} \\
&= \frac{(U^T(x_{j+1})[I - U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)]U(x))^2}{U^T(x_{j+1})[I - U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)]U(x_{j+1})},
\end{aligned}$$

where $[U^T(X_j)U(X_j)]^{-}$ denotes a generalized inverse of $[U^T(X_j)U(X_j)]$.

Let $C_{X_j}(t) = [I - U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)]U(t)$. Then,

$$C^T_{X_j}(x_{j+1})C_{X_j}(x) = U^T(x_{j+1})[I - U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)]U(x).$$

Similarly,

$$C^T_{X_j}(x_{j+1})C_{X_j}(x_{j+1}) = U^T(x_{j+1})[I - U(X_j)[U^T(X_j)U(X_j)]^{-}U^T(X_j)]U(x_{j+1}).$$

Therefore,

$$R(x_{n+1}) \approx \frac{(C_{X_j}^T(x_{j+1})C_{X_j}(x))^2}{C_{X_j}^T(x_{j+1})C_{X_j}(x_{j+1})} = \|C_{X_j}(x)\|_2^2 \cos^2(\vartheta),$$

where $\vartheta$ is the angle between $C_{X_j}(x)$ and $C_{X_j}(x_{j+1})$.

# Bibliography

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*, volume 157. Springer.