# Online Supplement for "Gaussian Process Based Optimization Algorithms With Input Uncertainty" by Haowei Wang, Jun Yuan & Szu Hui Ng

## A The proof of the Proposition 4.1

With the Dirichlet($\boldsymbol{\alpha} = \mathbf{1}$) prior and the categorical likelihood, the posterior is Dirichlet($\boldsymbol{\alpha'}$) with $\boldsymbol{\alpha'} = \mathbf{1} + \sum_{i=1}^{h} \boldsymbol{\xi_i}$, i.e. $p(\boldsymbol{\lambda}|\boldsymbol{D_h}) = \frac{\Pi_{k=1}^{K} \lambda_k^{\alpha'_k - 1}}{\mathrm{B}(\boldsymbol{\alpha'})}$. Noted that for $\forall k$, $\alpha'_k - 1 \geq 0$ and $\alpha'_k \in \mathbb{Z}$.

$$m_n(\boldsymbol{x}, \boldsymbol{\lambda}) = \tau^2 R_F((\boldsymbol{x}, \boldsymbol{\lambda}), \cdot; \boldsymbol{\theta})^T [\tau^2 R_F(\boldsymbol{\theta}) + \boldsymbol{\Sigma_\epsilon}]^{-1} \bar{\boldsymbol{Y}}_{\boldsymbol{n}} = \sum_{i=1}^{n} C_i(\boldsymbol{x}) \exp\left(\frac{-|\boldsymbol{\lambda} - \boldsymbol{\lambda_i}|}{2\theta_2^2}\right) = \sum_{i=1}^{n} C_i(\boldsymbol{x}) \Pi_{k=1}^{K} \exp\left(\frac{-|\lambda_k - \lambda_{i,k}|}{2\theta_{2,k}^2}\right)$$

where $C_i(\boldsymbol{x}) = \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x_i}|}{2\boldsymbol{\theta_1^2}}\right) \left(\tau^2 [\tau^2 R_F(\boldsymbol{\theta}) + \boldsymbol{\Sigma_\epsilon}]^{-1} \bar{\boldsymbol{Y}}_{\boldsymbol{n}}\right)_i$. Then

$$\mathrm{E}[G_n(\boldsymbol{x})] = \int_{\boldsymbol{\lambda}} m_n(\boldsymbol{x}, \boldsymbol{\lambda}) p(\boldsymbol{\lambda}|\boldsymbol{D_h}) d\boldsymbol{\lambda} = \frac{1}{B(\boldsymbol{\alpha'})} \sum_{i=1}^{n} C_i(\boldsymbol{x}) \Pi_{k=1}^{K} \int_{\lambda_k} \exp\left(\frac{-|\lambda_k - \lambda_{i,k}|}{2\theta_{2,k}^2}\right) \lambda_k^{\alpha'_k - 1} d\lambda_k \tag{1}$$

In order to compute (1) analytically, it is sufficient to derive the general analytical form of $\int_0^1 \exp\left(-\frac{|\lambda - \lambda_{i,k}|}{2\theta_{2,k}^2}\right) \lambda^n d\lambda$, $n \geq 0$. Recall that

$$\int_0^{\lambda} \exp(c\lambda) \lambda^n d\lambda = e^{c\lambda} \sum_{p=0}^{n} (-1)^{n-p} \frac{n!}{p! c^{n-p+1}} \lambda^p \quad \text{for } c \neq 0 \text{ and } n \geq 0 \tag{2}$$

$$\begin{aligned}
\int_0^1 \exp\left(\frac{-|\lambda - \lambda_{i,k}|}{2\theta_{2,k}^2}\right) \lambda^n d\lambda &= \int_0^{\lambda_{i,k}} \exp\left(\frac{\lambda - \lambda_{i,k}}{2\theta_{2,k}^2}\right) \lambda^n d\lambda + \int_{\lambda_{i,k}}^1 \exp\left(\frac{\lambda_{i,k} - \lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda \\
&= \exp\left(\frac{-\lambda_{i,k}}{2\theta_{2,k}^2}\right) \int_0^{\lambda_{i,k}} \exp\left(\frac{\lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda + \exp\left(\frac{\lambda_{i,k}}{2\theta_{2,k}^2}\right) \int_{\lambda_{i,k}}^1 \exp\left(\frac{-\lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda \\
&= \exp\left(\frac{-\lambda_{i,k}}{2\theta_{2,k}^2}\right) \int_0^{\lambda_{i,k}} \exp\left(\frac{\lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda + \exp\left(\frac{\lambda_{i,k}}{2\theta_{2,k}^2}\right) \left\{ \int_0^1 \exp\left(\frac{-\lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda - \int_0^{\lambda_{i,k}} \exp\left(\frac{-\lambda}{2\theta_{2,k}^2}\right) \lambda^n d\lambda \right\}
\end{aligned} \tag{3}$$

Equation (3) can be calculated analytically with the result in Equation (2). In addition,

$$\begin{aligned}
k_n((\boldsymbol{x}, \boldsymbol{\lambda}), (\boldsymbol{x'}, \boldsymbol{\lambda'})) &= \tau^2 \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x'}|}{2\boldsymbol{\theta_1^2}} - \frac{|\boldsymbol{\lambda} - \boldsymbol{\lambda'}|}{2\boldsymbol{\theta_2^2}}\right) - \sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} \exp\left(-\frac{|\boldsymbol{x} - \boldsymbol{x_i}|}{2\boldsymbol{\theta_1^2}} - \frac{|\boldsymbol{\lambda} - \boldsymbol{\lambda_i}|}{2\boldsymbol{\theta_2^2}} - \frac{|\boldsymbol{x'} - \boldsymbol{x_j}|}{2\boldsymbol{\theta_1^2}} - \frac{|\boldsymbol{\lambda'} - \boldsymbol{\lambda_j}|}{2\boldsymbol{\theta_2^2}}\right) \\
&= \tau^2 \exp\left(\frac{-|\boldsymbol{x} - \boldsymbol{x'}|}{2\boldsymbol{\theta_1^2}}\right) \Pi_{k=1}^{K} \exp\left(\frac{-|\lambda_k - \lambda'_k|}{2\theta_{2,k}^2}\right) - \sum_{j=1}^{n} \sum_{i=1}^{n} c_{ij} \exp\left(\frac{-|\boldsymbol{x} - \boldsymbol{x_i}| - |\boldsymbol{x'} - \boldsymbol{x_j}|}{2\boldsymbol{\theta_1^2}}\right) \Pi_{k=1}^{K} \left\{ \exp\left(-\frac{|\lambda_k - \lambda_{i,k}|}{2\theta_{2,k}^2}\right) \exp\left(-\frac{|\lambda'_k - \lambda_{j,k}|}{2\theta_{2,k}^2}\right) \right\}
\end{aligned}$$

where $c_{ij} = (\tau^4[\tau^2 R_F(\boldsymbol{\theta}) + \boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}]^{-1})_{ij}$. Then

$$\begin{aligned}
\mathrm{Cov}[G_n(\boldsymbol{x}), G_n(\boldsymbol{x'})] &= \int_{\boldsymbol{\lambda}}\int_{\boldsymbol{\lambda'}} p(\boldsymbol{\lambda}|\boldsymbol{D_h})p(\boldsymbol{\lambda'}|\boldsymbol{D_h})k_n((\boldsymbol{x},\boldsymbol{\lambda}),(\boldsymbol{x'},\boldsymbol{\lambda'}))d\boldsymbol{\lambda'}d\boldsymbol{\lambda} \\
&= \frac{\tau^2}{B(\boldsymbol{\alpha'})^2}\exp\left(\frac{-|\boldsymbol{x}-\boldsymbol{x'}|}{2\boldsymbol{\theta}_{\boldsymbol{1}}^2}\right)\Pi_{k=1}^K \int_{\lambda_k}\int_{\lambda'_k} {\lambda_k}^{\alpha'_k-1}{\lambda'_k}^{\alpha'_k-1}\exp\left(\frac{-|\lambda_k-\lambda'_k|}{2\theta_{2,k}^2}\right)d\lambda_k d\lambda'_k - \frac{1}{B(\boldsymbol{\alpha'})^2}\sum_{j=1}^n\sum_{i=1}^n c_{ij}\exp\left(\frac{-|\boldsymbol{x}-\boldsymbol{x_i}|-|\boldsymbol{x'}-\boldsymbol{x_j}|}{2\boldsymbol{\theta}_{\boldsymbol{1}}^2}\right) \\
&\cdot \Pi_{k=1}^K\left\{\int_{\lambda_k}\exp\left(-\frac{|\lambda_k-\lambda_{i,k}|}{2\theta_{2,k}^2}\right){\lambda_k}^{\alpha'_k-1}d\lambda_k\int_{\lambda'_k}\exp\left(-\frac{|\lambda'_k-\lambda_{j,k}|}{2\theta_{2,k}^2}\right){\lambda'_k}^{\alpha'_k-1}d\lambda'_k\right\}
\end{aligned} \tag{4}$$

In order to calculate (4), it is sufficient to show that $\int_0^1\int_0^1 \lambda^n\lambda'^m\exp\left(-\frac{|\lambda-\lambda'|}{2\theta_{2,k}^2}\right)d\lambda' d\lambda, n\geq 0, m\geq 0$ can be calculated analytically. In fact,

$$\begin{aligned}
\int_0^1\int_0^1 \lambda^n\lambda'^m\exp\left(-\frac{|\lambda-\lambda'|}{2\theta_{2,k}^2}\right)d\lambda' d\lambda &= \int_0^1\int_0^\lambda \lambda^n\lambda'^m\exp\left(\frac{-\lambda+\lambda'}{2\theta_{2,k}^2}\right)d\lambda' d\lambda + \int_0^1\int_\lambda^1 \lambda^n\lambda'^m\exp\left(\frac{-\lambda'+\lambda}{2\theta_{2,k}^2}\right)d\lambda' d\lambda \\
&= \int_0^1 \lambda^n\exp\left(\frac{-\lambda}{2\theta_{2,k}^2}\right)\int_0^\lambda \lambda'^m\exp\left(\frac{\lambda'}{2\theta_{2,k}^2}\right)d\lambda' d\lambda + \int_0^1 \lambda^n\exp\left(\frac{\lambda}{2\theta_{2,k}^2}\right)\int_\lambda^1 \lambda'^m\exp\left(\frac{-\lambda'}{2\theta_{2,k}^2}\right)d\lambda' d\lambda \\
&= \int_0^1 \lambda^n\exp\left(\frac{-\lambda}{2\theta_{2,k}^2}\right)\int_0^\lambda \lambda'^m\exp\left(\frac{\lambda'}{2\theta_{2,k}^2}\right)d\lambda' d\lambda + \int_0^1\left\{\lambda^n\exp\left(\frac{\lambda}{2\theta_{2,k}^2}\right)\left[\int_0^1 \lambda'^m\exp\left(\frac{-\lambda'}{2\theta_{2,k}^2}\right)d\lambda' - \int_0^\lambda \lambda'^m\exp\left(\frac{-\lambda'}{2\theta_{2,k}^2}\right)d\lambda'\right]\right\}d\lambda
\end{aligned} \tag{5}$$

can be calculated analytically using the result in (2) several times.

# B An $(s, S)$ inventory problem

In this section, we will focus on an $(s, S)$ inventory problem introduced in Example 1.1. For this problem, the decision makers need to decide the basic order level $s$ and the order-up-to level $S$ for a product. The random demand for this product is assumed to follow $\exp(\boldsymbol{\lambda})$. We adopt the analytical form $f(s, S, \boldsymbol{\lambda}) = \frac{1}{\boldsymbol{\lambda}} + [100 + s - \frac{1}{\boldsymbol{\lambda}} + 0.5\boldsymbol{\lambda}(S^2 - s^2) + \frac{101}{\boldsymbol{\lambda}}e^{-\boldsymbol{\lambda}s}]/[1 + \boldsymbol{\lambda}(S - s)]$ as the expected cost function (Jalali et al., 2017). In Jalali et al. (2017), they assumed that true input parameter is known ($\boldsymbol{\lambda} = 0.0002$), and solved the problem $\min_{s,S} f(s, S, 0.0002)$ via GP-based optimization algorithms. Here we consider the case where the true value of $\boldsymbol{\lambda}$ is unknown, and we have $h$ historical demand data: $\xi_1, \cdots, \xi_h \overset{i.i.d.}{\sim} \exp(\boldsymbol{\lambda} = 0.0002)$. We use the non-informative Jeffreys prior for $\boldsymbol{\lambda}$, then the posterior distribution for $\boldsymbol{\lambda}$ is $\mathrm{Gamma}(h, \sum_{i=1}^h \xi_i)$. The objective function is $g(s, S) = \mathrm{E}_{\boldsymbol{\lambda}}[f(s, S, \boldsymbol{\lambda})]$. The same algorithms in Section 5 are applied under the four scenarios with light (heavy) stochastic noise level and low (high) input uncertainty level. For this problem, we assume 100 and 10000 observations of real world data are available for high and low input uncertainty case respectively. The box-plots of GAP are shown in Figure 1.

Similar to the results from Section 5, all the algorithms perform similar with low input uncertainty. When input uncertainty is high, algorithms with PI perform very poorly while the algorithms with IMSE perform best. Algorithms with IMSE can be more advantageous over algorithms with

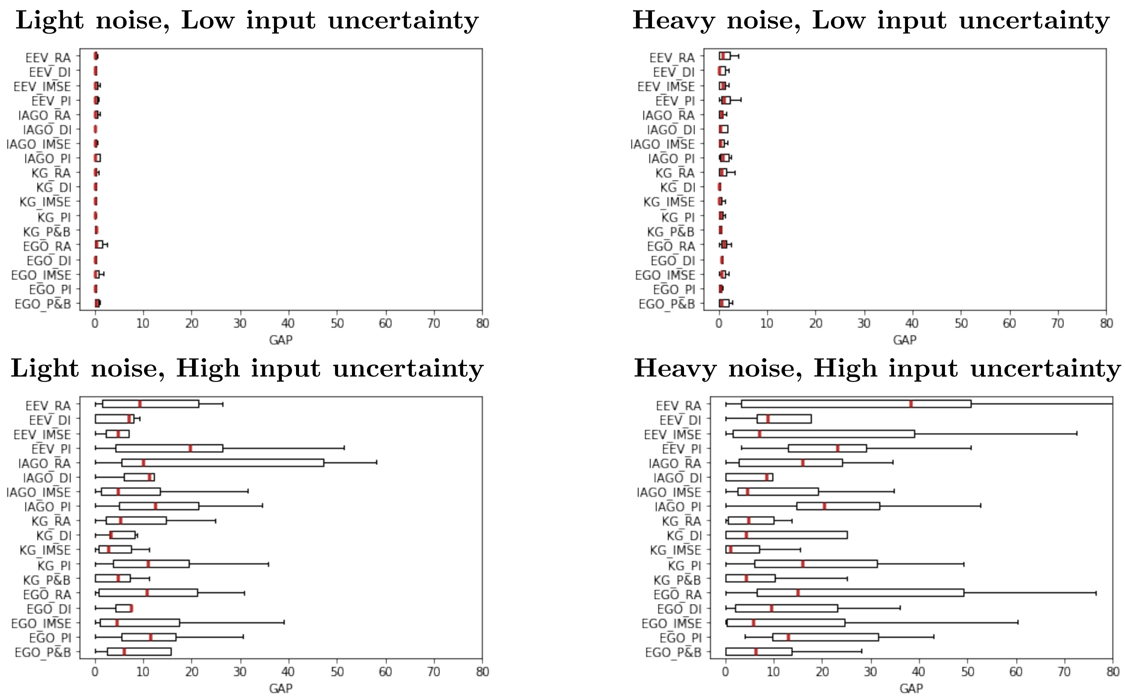


Figure 1: GAP for $(s, S)$ problem for different algorithms. The bold lines inside the boxes are the median values; the edges of the boxes are the 25th and 75th percentiles

PI when solving more realistic problems as it will be hard to measure the level of input uncertainty.

# C A correlated input parameter problem

In this section, we consider a 2-dim correlated input parameter problem using the Ackley-3 function. Specifically, we choose one of variable of the function to be the 1-dim design parameter $\boldsymbol{x}$, and treat another two variables as the 2-dim input parameter $\boldsymbol{\lambda} = [\lambda_1, \lambda_2]^T$ with the form: $f(\boldsymbol{x}, \boldsymbol{\lambda}) = -20 \exp\left\{-0.2\sqrt{\frac{1}{3}\left(\tilde{x}^2 + \tilde{\lambda}_1^2 + \tilde{\lambda}_2^2\right)}\right\} - \exp\left\{\frac{1}{3}\left[\cos(2\pi\tilde{x}) + \cos(2\pi\tilde{\lambda}_1) + \cos(2\pi\tilde{\lambda}_2)\right]\right\} + 20 + \exp(1)$, where $\tilde{x} = 10x, \tilde{\lambda}_1 = 10\lambda_1, \tilde{\lambda}_2 = 10\lambda_2, x \in [-3, 3]$. We assume the real world data $\xi_1, \cdots, \xi_h \overset{i.i.d.}{\sim} N(\lambda_1, \lambda_2)$, where $\lambda_1$ is the unknown mean and $\lambda_2$ is the unknown variance. The true value of $\lambda_1$ is 1 and true value of $\lambda_2$ is 1. We use the non-informative Jeffreys prior for the unknown mean and unknown variance: $P(\lambda_1, \lambda_2) \propto 1/\lambda_2$. Then the posterior is Normal-Inverse-Gamma distribution and direct sampling from the posterior is possible. The same algorithms in Section 5 are applied under the four scenarios with light (heavy) stochastic noise level and low (high) input uncertainty level. For this problem, we assume 10 and 100 observations of real world data are available for high

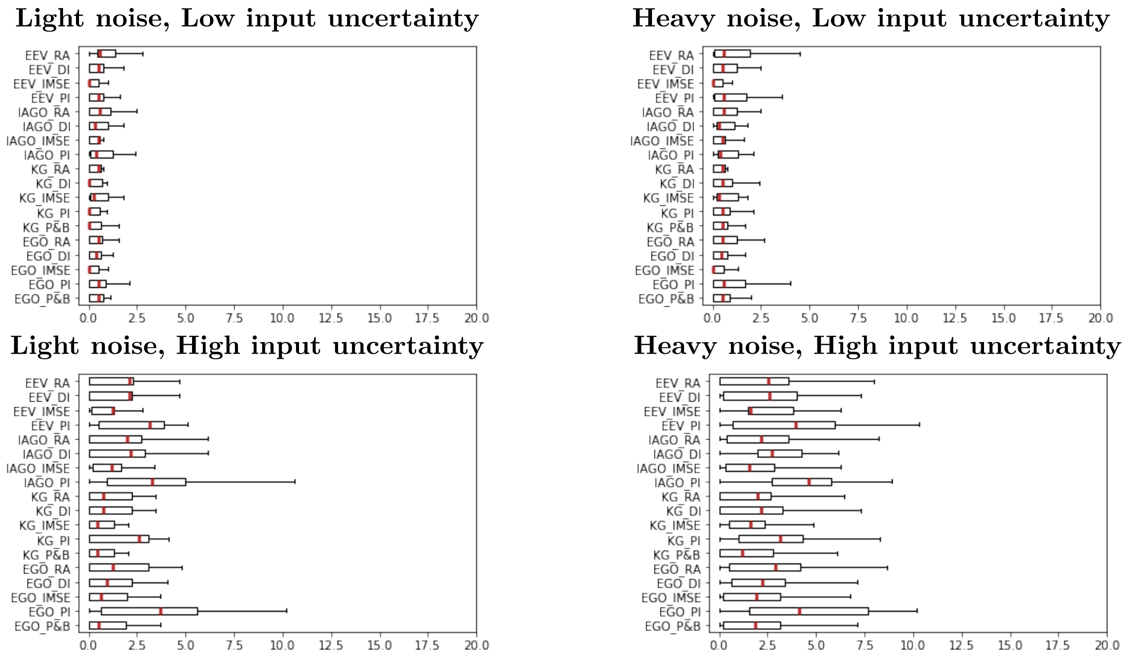


Figure 2: GAP for the correlated input parameter problem for different algorithms. The bold lines inside the boxes are the median values; the edges of the boxes are the 25th and 75th percentiles

and low input uncertainty case respectively. The box-plots of GAP are shown in Figure 2.

The figure shows that with high input uncertainty, the algorithms with PI perform worse than algorithms with input uncertainty, and the algorithms with IMSE perform consistently much better than algorithms with DI and RA with the correlated input parameter problem. This may be due to that, in the correlated case, the DI and RA will tend to evaluate the region where the joint density of the different dimensions of the input parameter is high and not be able to cover the whole region. In general, the results for the correlated input parameter are similar to the results in Section 5.

# References

Jalali, H., Van Nieuwenhuyse, I. and Picheny, V. (2017). Comparison of kriging-based algorithms for simulation optimization with heterogeneous noise, *European Journal of Operational Research* **261**(1): 279–301.