A. Additional Background

A.1. Sobolev Spaces

The RKHS induced from the Matérn kernel c_α defined in Equation 1 is norm-equivalent to a Sobolev space (Adams & Fournier, 2003). When $\alpha \in \mathbb{N}$, these spaces are defined as:

$$W_2^\alpha(\mathcal{X}) \quad := \quad \{f \in L_2(\mathcal{X}) : D^\nu f \in L_2(\mathcal{X}) \text{ exists } \forall \nu \in \mathbb{N}_0^p \text{ with } |\nu| \leq \alpha\},$$

with inner product

$$\langle f, g \rangle_{W_2^\alpha(\mathcal{X})} \quad := \quad \sum_{|\nu| \leq \alpha} \langle D^\nu f, D^\nu g \rangle_{L_2(\mathcal{X})}$$

for all $f, g \in W_2^\alpha(\mathcal{X})$. This means that all functions in $\mathcal{H}_{k_\alpha}$ will have smoothness α (here D^ν denotes the total derivative corresponding to the multi-index $\nu = (\nu_1, \ldots, \nu_p) \in \mathbb{N}_0^p$).

It is also possible to have fractional Sobolev spaces; i.e. the smoothness $\alpha > 0$ can take any positive real value. For $\mathcal{X} = \mathbb{R}^p$ and denoting by $\hat{f}$ the Fourier transform of f, these spaces are given by:

$$H^\alpha(\mathbb{R}^p) \quad := \quad \left\{ f \in L_2(\mathbb{R}^p) : \int |\hat{f}(\xi)|^2 (1 + \|\xi\|^2)^\alpha \mathrm{d}\xi < \infty \right\}$$

with associated inner product:

$$\langle f, g \rangle_{H^\alpha(\mathbb{R}^p)} \quad := \quad \int \hat{f}(\xi) \overline{\hat{g}(\xi)} (1 + \|\xi\|^2)^\alpha \mathrm{d}\xi$$

for all $f, g \in H^\alpha(\mathbb{R}^d)$ where $\overline{\hat{g}}$ denoted the complex conjugate of $\hat{g}$.

A.2. Lipschitz Boundary Conditions

In this section we introduce the notion of Lipschitz boundary condition, which is required for our domain $\mathcal{X}$ in the theory in Section 3. The introduction in this section follows that of Section 3 in (Kanagawa et al., 2017).

To do so, we begin by introducing *special Lispchitz domains*. For $d > 2$, we say that an open set $\mathcal{X} \subset \mathbb{R}^p$ is a special Lipschitz domain if there exists a rotation of $\mathcal{X}$, denoted by $\tilde{\mathcal{X}}$, and a function $\phi : \mathbb{R}^{p-1} \to \mathbb{R}$ that satisfy the following:

1. $\tilde{\mathcal{X}} = \{(x, y) \in \mathbb{R}^p : y > \phi(x)\}$.
2. ϕ is a Lipschitz function such that $|\phi(x) - \phi(x')| \leq M\|x - x'\| \forall x, x' \in \mathbb{R}^{p-1}$, where $M > 0$ is called the Lipschitz bound of $\mathcal{X}$.

With this definition now complete, we can define the notion of a domain with Lipschitz boundary. Let $\mathcal{X} \subset \mathbb{R}^p$ be an open set and $\partial\mathcal{X}$ be its boundary. We say the boundary is Lipschitz $\exists \epsilon, M > 0, K \in \mathbb{N}$ and open sets $U_1, \ldots, U_L \subset \mathbb{R}^p$ where $L \in \mathbb{N} \cup \{\infty\}$ such that the following holds:

1. For any $x \in \partial\mathcal{X}$, $\exists i$ such that $B(x, \epsilon)$, the ball centred at x of radius ϵ, satisfies $B(x, \epsilon) \subset U_i$.
2. $U_{i_1} \cap \ldots \cap U_{i_{K+1}} = \emptyset$ for any distinct indices $\{i_1, \ldots, i_{K+1}\}$.
3. For each index i, $\exists$ a special Lipschitz domain $\mathcal{X}_i \subset \mathbb{R}^p$ with Lipschitz bound b such that $U_i \cap \mathcal{X} = U_i \cap \mathcal{X}_i$ and $b \leq M$.

B. Proofs

PROOF OF PROPOSITION 1

Proof. This proof follows directly the proof for the uni-output case in (Briol et al., 2015b). Suppose we have a prior on $\boldsymbol{f}$, denoted $\boldsymbol{g}$, which is a Gaussian process $\mathcal{GP}(\boldsymbol{0}, \boldsymbol{C})$. Conditioning on some observations $(\boldsymbol{X}, \boldsymbol{Y}) = \{(\boldsymbol{X}_j, \boldsymbol{Y}_j)\}_{j=1}^D$, we get

a Gaussian process posterior $\boldsymbol{g}_N$ where the mean and covariance functions are given by:

$$\begin{aligned}\boldsymbol{m}_N(\boldsymbol{x}) &= \boldsymbol{C}(\boldsymbol{x},\boldsymbol{X})\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{X}),\\ \boldsymbol{C}_N(\boldsymbol{x},\boldsymbol{x}') &= \boldsymbol{C}(\boldsymbol{x},\boldsymbol{x}') - \boldsymbol{C}(\boldsymbol{x},\boldsymbol{X})\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{C}(\boldsymbol{X},\boldsymbol{x}').\end{aligned}$$

Several applications of Fubini's theorem on each element of the vectors give:

$$\begin{aligned}\mathbb{E}\left[\Pi[\boldsymbol{g}_N]\right] &= \int_\Omega\int_{\mathcal{X}} \boldsymbol{g}_N(\boldsymbol{x},\omega)\Pi(\mathrm{d}\boldsymbol{x})\mathbb{P}(\mathrm{d}\omega) = \int_{\mathcal{X}} \boldsymbol{m}_N(\boldsymbol{x})\Pi(\mathrm{d}\boldsymbol{x}) = \Pi[\boldsymbol{C}(\cdot,\boldsymbol{X})]\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{X}),\\ \mathbb{V}[\Pi[\boldsymbol{g}_N]] &= \int_\Omega\left[\int_{\mathcal{X}} \boldsymbol{g}_N(\boldsymbol{x},\omega)\Pi(\mathrm{d}\boldsymbol{x}) - \int_{\mathcal{X}} \boldsymbol{m}_N(\boldsymbol{x})\Pi(\mathrm{d}\boldsymbol{x})\right]^2\mathbb{P}(\mathrm{d}\omega)\\ &= \int_{\mathcal{X}}\int_{\mathcal{X}}\int_\Omega \left[\boldsymbol{g}_N(\boldsymbol{x},\omega)-\boldsymbol{m}_N(\boldsymbol{x})\right]\left[\boldsymbol{g}_N(\boldsymbol{x}',\omega)-\boldsymbol{m}_N(\boldsymbol{x}')\right]\mathbb{P}(\mathrm{d}\omega)\Pi(\mathrm{d}\boldsymbol{x})\Pi(\mathrm{d}\boldsymbol{x}')\\ &= \int_{\mathcal{X}}\int_{\mathcal{X}} \boldsymbol{C}_N(\boldsymbol{x},\boldsymbol{x}')\Pi(\mathrm{d}\boldsymbol{x})\Pi(\mathrm{d}\boldsymbol{x}')\\ &= \Pi\bar{\Pi}[\boldsymbol{C}] - \Pi[\boldsymbol{C}(\cdot,\boldsymbol{X})]\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\bar{\Pi}[\boldsymbol{C}(\boldsymbol{X},\cdot)].\end{aligned}$$

□

PROOF OF PROPOSITION 2

Proof. Denote by $\boldsymbol{e}_d$ the vertical vector of length D with d^{th} entry taking value 1 and all other entries taking value 0, and by $\boldsymbol{C}^d_{\boldsymbol{x}}(\boldsymbol{y}) = \boldsymbol{C}(\boldsymbol{y},\boldsymbol{x})\boldsymbol{e}_d$ the d^{th} column of $\boldsymbol{C}(\boldsymbol{y},\boldsymbol{x})$. We notice that the representer of the integral is given by:

$$\Pi[f_d] = \Pi[\boldsymbol{f}^\top\boldsymbol{e}_d] = \Pi\left[\langle \boldsymbol{f}, \boldsymbol{C}(\cdot,\boldsymbol{x})\boldsymbol{e}_d\rangle_{\boldsymbol{C}}\right] = \left\langle \boldsymbol{f}, \Pi\left[\boldsymbol{C}(\cdot,\boldsymbol{x})\boldsymbol{e}_d\right]\right\rangle_{\boldsymbol{C}} = \left\langle \boldsymbol{f}, \Pi\left[\boldsymbol{C}^d_{\boldsymbol{x}}\right]\right\rangle_{\boldsymbol{C}}.$$

Using the Cauchy-Schwartz inequality, we get:

$$\left|\Pi[f_d] - \hat{\Pi}[f_d]\right| \leq \|\boldsymbol{f}\|_{\boldsymbol{C}}\left\|\Pi[\boldsymbol{C}^d_{\boldsymbol{x}}] - \hat{\Pi}[\boldsymbol{C}^d_{\boldsymbol{x}}]\right\|_{\boldsymbol{C}}.$$

Taking supremums, we then obtain the following expression for the worst-case integration error:

$$\sup_{\|f\|_{\boldsymbol{C}}\leq 1}\left|\Pi[f_d] - \hat{\Pi}[f_d]\right| = \left\|\Pi[\boldsymbol{C}^d_{\boldsymbol{x}}] - \hat{\Pi}[\boldsymbol{C}^d_{\boldsymbol{x}}]\right\|_{\boldsymbol{C}}.$$

We note that $\Pi[\boldsymbol{C}^d_{\boldsymbol{x}}] \in \mathcal{H}_{\boldsymbol{C}}$ and that the multi-output BQ rule is given by

$$\hat{\Pi}_{\text{BQ}}[\boldsymbol{C}^d_{\boldsymbol{x}}] = \Pi[\boldsymbol{C}(\cdot,\boldsymbol{X})]\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{C}^d_{\boldsymbol{x}}(\boldsymbol{X}),$$

and corresponds to an optimal interpolant in the sense of Thm 3.1 (Micchelli & Pontil, 2005). We must therefore have that, for fixed quadrature points $\boldsymbol{X}$, any quadrature rule $\hat{\Pi}[\boldsymbol{C}^d_{\boldsymbol{x}}]$ satisfies:

$$\left\|\Pi[\boldsymbol{C}^d_{\boldsymbol{x}}] - \hat{\Pi}_{\text{BQ}}[\boldsymbol{C}^d_{\boldsymbol{x}}]\right\|_{\boldsymbol{C}} \leq \left\|\Pi[\boldsymbol{C}^d_{\boldsymbol{x}}] - \hat{\Pi}[\boldsymbol{C}^d_{\boldsymbol{x}}]\right\|_{\boldsymbol{C}}.$$

Combining the equation above with the expression for the worst-case integration error of f_d gives us our final result.

□

PROOF OF THEOREM 1

Proof. For the sake of clarity, we will distinguish between uni-output BQ and multi-output BQ rules and weights by adding subscripts corresponding to their kernel; i.e. $\Pi^{\boldsymbol{C}}_{\text{BQ}}[f]$ and $\boldsymbol{W}^{\boldsymbol{C}}_{\text{BQ}}$ denote the multi-output case and $\Pi^{c}_{\text{BQ}}[f]$ and $\boldsymbol{W}^{c}_{\text{BQ}}$ denote the uni-output case.

We start this proof by writing an expression for the weights of the multi-output BQ algorithm in terms of weights for the uni-output BQ algorithm:

$$\begin{aligned}
\boldsymbol{W}_{\mathrm{BQ}}^{\boldsymbol{C}} &= \left(\Pi[\boldsymbol{C}(\cdot,\boldsymbol{X})]\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\right)^{\top} \\
&= \left(\left(\Pi[\boldsymbol{B}\otimes\boldsymbol{c}(\cdot,\boldsymbol{X})]\right)\left(\boldsymbol{B}\otimes c(\boldsymbol{X},\boldsymbol{X})\right)^{-1}\right)^{\top} \\
&= \left(\left(\boldsymbol{B}\otimes\Pi[\boldsymbol{c}(\cdot,\boldsymbol{X})]\right)\left(\boldsymbol{B}^{-1}\otimes c(\boldsymbol{X},\boldsymbol{X})^{-1}\right)\right)^{\top} \\
&= \left(\boldsymbol{B}\boldsymbol{B}^{-1}\otimes\Pi[\boldsymbol{c}(\cdot,\boldsymbol{X})]c(\boldsymbol{X},\boldsymbol{X})^{-1}\right)^{\top} \\
&= \left(\boldsymbol{I}_D\otimes\left(\boldsymbol{w}_{\mathrm{BQ}}^{c}\right)^{\top}\right)^{\top} \\
&= \boldsymbol{I}_D\otimes\boldsymbol{w}_{\mathrm{BQ}}^{c}.
\end{aligned}$$

Using the above, we can find an expression for the multi-output BQ approximation with some kernel $\boldsymbol{C}_1 = \boldsymbol{B}c_1$ of the project mean element with respect to kernel $\boldsymbol{C}_2 = \boldsymbol{B}c_2$ in terms of the uni-output BQ approximation with kernel c_1 of the kernel mean of c_2.

$$\begin{aligned}
\hat{\Pi}_{\mathrm{BQ}}^{\boldsymbol{C}_1}[(\boldsymbol{C}_2)_{\boldsymbol{x}}^{d}] &= \left(\boldsymbol{W}_{\mathrm{BQ}}^{C_1}\right)^{\top}(\boldsymbol{C}_2)_{\boldsymbol{x}}^{d}(\boldsymbol{X}) \\
&= (I\otimes\boldsymbol{w}_{\mathrm{BQ}}^{c_1})^{\top}(\boldsymbol{C}_2)_{\boldsymbol{x}}^{d}(\boldsymbol{X}) \\
&= (\boldsymbol{I}\otimes\boldsymbol{w}_{\mathrm{BQ}}^{c_1})^{\top}(\boldsymbol{B}\boldsymbol{e}_d\otimes c_2(\boldsymbol{X},\boldsymbol{x})) \\
&= \boldsymbol{I}\boldsymbol{B}\boldsymbol{e}_d\otimes\left(\boldsymbol{w}_{\mathrm{BQ}}^{c_1}\right)^{\top}c_2(\boldsymbol{X},\boldsymbol{x}) \\
&= \boldsymbol{B}\boldsymbol{e}_d\hat{\Pi}_{\mathrm{BQ}}^{c_1}[c_2(\cdot,\boldsymbol{x})].
\end{aligned}$$

As discussed, taking both kernels to be the same, the integration error for each individual integrand can be bounded as follows:

$$\begin{aligned}
\sup_{\|f\|_{\boldsymbol{C}_2}\leq 1}\left|\Pi[f_d]-\hat{\Pi}_{\mathrm{BQ}}^{\boldsymbol{C}_1}[f_d]\right|^2 &= \left\|\Pi\left[(\boldsymbol{C}_2)_{\boldsymbol{x}}^{d}\right]-\hat{\Pi}_{\mathrm{BQ}}^{\boldsymbol{C}_1}\left[(\boldsymbol{C}_2)_{\boldsymbol{x}}^{d}\right]\right\|_{\boldsymbol{C}_2}^2 \\
&= \left\|(\boldsymbol{B}\boldsymbol{e}_d)\left(\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right]\right)\right\|_{\boldsymbol{C}_2}^2 \\
&= \sum_{i,j=1}^{D}(\boldsymbol{B}^{-1})_{ij}\times\left\langle\boldsymbol{B}_{id}(\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right]),\boldsymbol{B}_{jd}(\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right])\right\rangle_{c_2} \\
&= \sum_{i,j=1}^{D}(\boldsymbol{B}^{-1})_{ij}\boldsymbol{B}_{id}\boldsymbol{B}_{jd}\left\|\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right]\right\|_{c_2}^2 \\
&\leq K\left\|\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right]\right\|_{c_2}^2.
\end{aligned}$$

Here, we first used the definition of worst-case error, then the definition of the $H_{\boldsymbol{C}_2}$ norm in terms of $\mathcal{H}_{c_2}$ norm (as given for the seperable kernel in (Alvarez et al., 2012)), and the final inequality follows by taking $K > 0$ to be $K = |\sum_{i,j=1}^{D}(\boldsymbol{B}^{-1})_{ij}\boldsymbol{B}_{id}\boldsymbol{B}_{jd}|$. Taking the square-root on either side gives us:

$$\sup_{\|f\|_{\boldsymbol{C}_2}\leq 1}\left|\Pi[f_d]-\hat{\Pi}_{\mathrm{BQ}}^{\boldsymbol{C}_1}[f_d]\right| \leq \sqrt{K}\left\|\Pi\left[c_2(\cdot,\boldsymbol{x})\right]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}\left[c_2(\cdot,\boldsymbol{x})\right]\right\|_{c_2} = \sqrt{K}\sup_{\|f\|_{c_2}\leq 1}\left|\Pi[f_d]-\hat{\Pi}_{\mathrm{BQ}}^{c_1}[f_d]\right|. \quad (3)$$

We can take $\boldsymbol{C}_1$ equal to $\boldsymbol{C}_2$ to get:

$$\sup_{\|f\|_{\boldsymbol{C}}\leq 1}\left|\Pi[f_d]-\hat{\Pi}_{\mathrm{BQ}}^{\boldsymbol{C}}[f_d]\right| \leq \sqrt{K}\sup_{\|f\|_{c}\leq 1}\left|\Pi[f_d]-\hat{\Pi}_{\mathrm{BQ}}^{c}[f_d]\right|.$$

The convergence for the separable kernel case is therefore driven by the convergence of the scalar-valued kernel. We can therefore use results from the uni-output case in (Briol et al., 2015b; Oates et al., 2016; Briol et al., 2017; Kanagawa et al., 2017) to complete the proof. □

PROOF OF PROPOSITION 3

Proof. Note that if the kernel is actually of the form $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{x}') = \sum_{q=1}^{Q} \boldsymbol{B}_q c_q(\boldsymbol{x}, \boldsymbol{x}')$, we can use the triangle inequality satisfied by the norm of $\mathcal{H}_{\boldsymbol{C}}$ to show that (for some $C_2 > 0$):

$$\sup_{\|f\|_{\boldsymbol{C}} \leq 1} \left| \Pi[f_d] - \hat{\Pi}_{\text{BQ}}[f_d] \right| \quad \leq \quad C_2 \sum_{q=1}^{Q} \left\| \Pi\left[c_q(\cdot, \boldsymbol{x})\right] - \hat{\Pi}_{\text{BQ}}\left[c_q(\cdot, \boldsymbol{x})\right] \right\|_c^2,$$

so that the overall convergence is dominated by the slowest decaying term. □

B.1. Proof of Theorem 2

Proof. Denote by $\hat{\Pi}_{BQ}^{\boldsymbol{C}_\alpha}[\boldsymbol{f}]$ the multi-output BQ rule based on $\boldsymbol{C}_\alpha$, $\hat{\Pi}_{BQ}^{c_\alpha}[f]$ the uni-output BQ rule based on c_α and $\hat{f}_d^\alpha$ the interpolant corresponding this rule. We start by upper bounding the integration error in the uni-output case:

$$\begin{aligned}
\left| \Pi[f] - \hat{\Pi}_{BQ}^{c_\alpha}[f] \right| &\leq K_1 \|\pi\|_{L_\infty(\mathcal{X})} \|f - \hat{f}^\alpha\|_{L_1(\mathcal{X})} \\
&\leq K_2 \|f - \hat{f}^\alpha\|_{L_2(\mathcal{X})} \\
&\leq K_3 h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha \|f\|_{L_2(\mathcal{X})} \\
&\leq K_4 h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha \|f\|_{W_2^\beta(\mathcal{X})} \\
&\leq K_5 h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha \|f\|_{c_\beta},
\end{aligned}$$

for some $K_1, \ldots, K_5 > 0$. The first and second inequality correspond to Holder's inequality and the third inequality follows from Theorem 4.2 in (Narcowich et al., 2006). Finally, the fourth and fifth inequalities follow from the definition the Sobolev norm and the norm-equivalence of $\mathcal{H}_{c_\beta}$ and $W_2^\beta(\mathcal{X})$.

Dividing the above by $\|f_d\|_\beta$ on both sides and taking supremums over the unit ball of $\mathcal{H}_{c_\beta}$ we get a result for the worst-case error in the uni-output case:

$$e(\mathcal{H}_{c_\beta}, \hat{\Pi}_{BQ}^{c_\alpha}, \boldsymbol{X}) \leq K_6 h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha.$$

We can then upper bound the integration error in the multi-output case using Theorem 1 as follows:

$$\begin{aligned}
\left| \Pi[f_d] - \hat{\Pi}_{BQ}^{\boldsymbol{C}_\alpha}[f_d] \right| &\leq \|\boldsymbol{f}\|_{\boldsymbol{C}_\beta} e(\mathcal{H}_{\boldsymbol{C}_\beta}, \hat{\Pi}_{BQ}^{\boldsymbol{C}_\alpha}, \boldsymbol{X}, d) \\
&\leq K_6 \|\boldsymbol{f}\|_{\boldsymbol{C}_\beta} e(\mathcal{H}_{c_\beta}, \hat{\Pi}_{BQ}^{c_\alpha}, \boldsymbol{X}) \\
&\leq K_7 \|\boldsymbol{f}\|_{\boldsymbol{C}_\beta} h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha,
\end{aligned}$$

for some $K_6, K_7 > 0$. When (A_2) is satisfied, then we can use the assumption that $h_{\boldsymbol{X},\mathcal{X}} \leq C q_{\boldsymbol{X}}$ for some constant $C > 0$ and the fact that $h_{\boldsymbol{X},\mathcal{X}}$ converges as $N^{-\frac{1}{p}}$ to show that the integration error satisfies:

$$\left| \Pi[f_d] - \hat{\Pi}_{BQ}^{\boldsymbol{C}_\alpha}[f_d] \right| \quad \leq \quad K_7 \|\boldsymbol{f}\|_{\boldsymbol{C}_\beta} h_{\boldsymbol{X},\mathcal{X}}^\beta \rho_{\boldsymbol{X},\mathcal{X}}^\alpha \;\leq\; K_8 \|\boldsymbol{f}\|_{\boldsymbol{C}_\beta} h_{\boldsymbol{X},\mathcal{X}}^\beta \;=\; \mathcal{O}\left(N^{-\frac{\beta}{p}}\right),$$

for some $K_8 > 0$. □

C. Implementation

In this appendix, we present some complementary details which will help users reproduce experiments in the paper.

C.1. Prior specification

C.1.1. SEPARABLE KERNEL

The separable matrix-valued kernel is of the form $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{x}') = \boldsymbol{B} c(\boldsymbol{x}, \boldsymbol{x}')$ where $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a scalar-valued kernel. If all of the elements f_d of the vector-valued function $\boldsymbol{f}$ are evaluated on the same data set $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$, then the Gram matrix can be expressed as

$$\boldsymbol{C}(\boldsymbol{X}, \boldsymbol{X}) \quad = \quad \boldsymbol{B} \otimes c(\boldsymbol{X}, \boldsymbol{X}),$$

where $\otimes$ denotes the Kronecker product. Due to properties of the Kronecker, its inverse can then be computed as:

$$\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1} \quad = \quad \boldsymbol{B}^{-1} \otimes c(\boldsymbol{X},\boldsymbol{X})^{-1}.$$

It is straightforward to show that similar expressions can be obtained for the multi-output case of the kernel mean:

$$\Pi[\boldsymbol{C}(\cdot,\boldsymbol{X})] \quad = \quad \boldsymbol{B} \otimes \Pi[c(\cdot,\boldsymbol{X})] \; = \; \boldsymbol{B} \otimes \left(\int_{\mathcal{X}} c(\boldsymbol{x},\boldsymbol{X}) \Pi(\mathrm{d}\boldsymbol{x}) \right),$$

and initial error:

$$\Pi\Pi[\boldsymbol{C}] \quad = \quad \boldsymbol{B} \; \Pi\bar{\Pi}[c] \; = \; \boldsymbol{B} \int_{\mathcal{X}\times\mathcal{X}} c(\boldsymbol{x},\boldsymbol{x}') \Pi(\mathrm{d}\boldsymbol{x}) \Pi(\mathrm{d}\boldsymbol{x}').$$

These expressions can of course be obtained in closed form whenever the kernel mean and initial error of the scalar-valued kernel are available in closed form. We refer the reader to the table in (Briol et al., 2015b) for a list of popular kernels for which this is possible.

C.1.2. Process Convolution kernel

In this section, we consider the process convolution kernel given by:

$$\begin{aligned}
(\boldsymbol{C}(x,x'))_{d,d'} &= c_{d,d'}(x,x') + c_{w_d}(x,x')\delta_{d,d'}, \\
c_{d,d'}(x,x') &= \sum_{i=1}^{R} \int_{\mathcal{X}} G_d^i(x-z) \int_{\mathcal{X}} G_{d'}^i(x'-z') c_i(z,z') \mathrm{d}z' \mathrm{d}z,
\end{aligned}$$

This is used in Sec. 4 in the two-output case. There, blurring kernels and reproducing kernels are:

$$\begin{aligned}
G_1^1(r) &= \lambda_1^2 \exp\left(-\frac{r^2}{2\sigma_1^2}\right), \\
G_2^1(r) &= \lambda_2^2 \exp\left(-\frac{r^2}{2\sigma_2^2}\right), \\
c_1(x,y) &= \lambda_3^2 \exp\left(-\frac{(x-y)^2}{2\sigma_3^2}\right), \\
G_1^2(r) &= \lambda_4^2 \exp\left(-\frac{r^2}{2\sigma_4^2}\right), \\
G_2^2(r) &= \lambda_5^2 \exp\left(-\frac{r^2}{2\sigma_5^2}\right), \\
c_2(x,y) &= \lambda_6^2 \exp\left(-\frac{(x-y)^2}{2\sigma_6^2}\right),
\end{aligned}$$

for some constants $\sigma_i, \lambda_i > 0$ for $i = 1, \ldots, 6$. Note that for simplicity, we did not include c_{w_1} and c_{w_2}. The kernel mean and initial error can easily be computed in closed form using Gaussian identities.

C.2. Hyper-parameters

One of the main challenges when using uni-output BQ and multi-output BQ is the selection of appropriate hyperparameters. In this section, we consider multi-output BQ with $\mathcal{GP}(\boldsymbol{0}, \boldsymbol{C})$ prior we denote the hyperparameters of the kernel $\boldsymbol{C}$ in vector form as $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_l)$. To optimise these parameters, we propose to use an empirical-Bayes approach and maximise the log-marginal likelihood:

$$\log p\left(\boldsymbol{f}(\boldsymbol{X})|\boldsymbol{X},\boldsymbol{\theta}\right) \quad = \quad -\frac{1}{2}\boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{X}) - \frac{1}{2}\log|\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})| - \frac{ND}{2}\log(2\pi).$$

This can be efficiently optimised by making use of gradients of the log-marginal likelihood $\forall i \in \{1, \ldots, l\}$:

$$\frac{\partial \log p\left(\boldsymbol{f}(\boldsymbol{X})|\boldsymbol{X},\boldsymbol{\theta}\right)}{\partial \theta_i} \quad = \quad \frac{1}{2}\boldsymbol{f}(\boldsymbol{X})^\top \boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1} \frac{\partial \boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})}{\partial \theta_i} \boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\boldsymbol{f}(\boldsymbol{X}) - \frac{1}{2}\mathrm{Tr}\left(\boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})^{-1}\frac{\partial \boldsymbol{C}(\boldsymbol{X},\boldsymbol{X})}{\partial \theta_i}\right).$$

D. Extended Numerical Experiments

In this appendix, we provide additional results for the multi-output BQ experiments provided in Sec. 4 for the multifidelity toy models and the global illumination problem. We also include numerical results for a popular variational approximation of multi-output GP.

D.1. Scaling multi-output BQ with variational approximations

Computational burdens are heavy for multi-output BQ due to the inversion of $ND \times ND$ matrix. The computational complexity is $O(N^3D^3)$ and the storage is $O(N^2D^2)$. (Álvarez & Lawrence, 2011) introduced and fully discussed a sparse approximation of multi-output GPs with process convolution kernels, using the fact that outputs are conditionally independent if the latent functions is fully observed. This idea can then be extended to multi-output BQ by taking our posterior on the value of the integrals as the pushforward through the integral operator of the approximate multi-output GP.

Consider functions $f_1(x) = 3\cos(\frac{\pi x}{5})$ and $f_2(x) = 0.7\cos(\frac{2.2\pi x}{10})$ on $\mathcal{X} = [-5, 5]$. Computational times and log-errors of multi-output BQ estimates for integrals of these functions against a uniform measure Π with and without the variational approximation by (Álvarez & Lawrence, 2011) using different number of equidistant points between -5 and 5 are given in Fig. 5. This approximation is considered for different number of points evaluated from the latent function, i.e. $C = 5, 15, 25, 35$. Regarding the process convolution kernel, G_1^1, G_1^2, c_1, G_2^1, G_2^2, c_2 are squared-exponential kernels with amplitude and lengthscale parameters $(\sqrt{3}, 1.3)$, $(0.7, 1)$, $(1, 1)$, $(0.9, 0.6)$, $(0.6, 0.5)$ and $(0.8, 1)$ respectively.

Clearly, with a large enough number of points C, the same integration accuracy as for the full GP can be obtained at much lower computation cost. This could make variational approximations a promising approach for multi-output BQ, but this would warrant a much more extensive study.

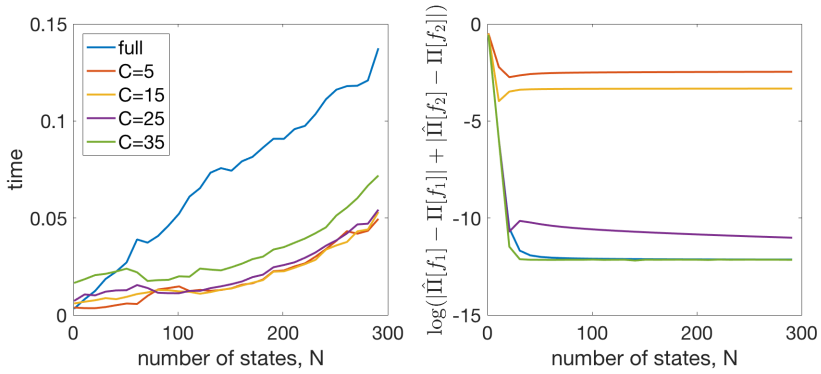

Figure 5. Variational approximation: Plot of the computational times (in seconds) and log-errors of full/ approximated multi-output BQ against the number of points given for different number of points evaluated from the latent function.

D.2. Multi-fidelity modelling

In Fig. 6, we give an extended version of Fig. 1 which includes credible intervals for each of the multi-output GP models. For both functions, the high-fidelity confidence intervals in the uni-output case are overly pessimistic, whereas for the multi-output cases, the posterior is concentrated on the true functions. One interesting point is that both of the multi-output BQ methods are over-confident near the kinks in the functions. This is to be expected since the true functions do not like in the RKHS corresponding to the kernel used for these BQ rules.

As an extension to these experiments, we consider a steady-state version of the Allen-Cahn equation on $[0, 10]$ subject to a sinusoidal forcing term and with boundary conditions:

$$\epsilon \frac{\partial^2 u}{\partial x^2} + u - u^3 = \sin(x), \quad u(0) = 1, \quad u(10) = -1, \tag{4}$$

where ϵ controls the rate of diffusion (Trefethen, 2010). Our target is to approximate the integral of the solution of Eq. 4 for $\epsilon \approx 0$. We take solutions of Eq. 4 on $\mathcal{X} = [0, 10]$ with $\epsilon = 2$ and $\epsilon = 0.1$ as our low-fidelity model and high-fidelity model

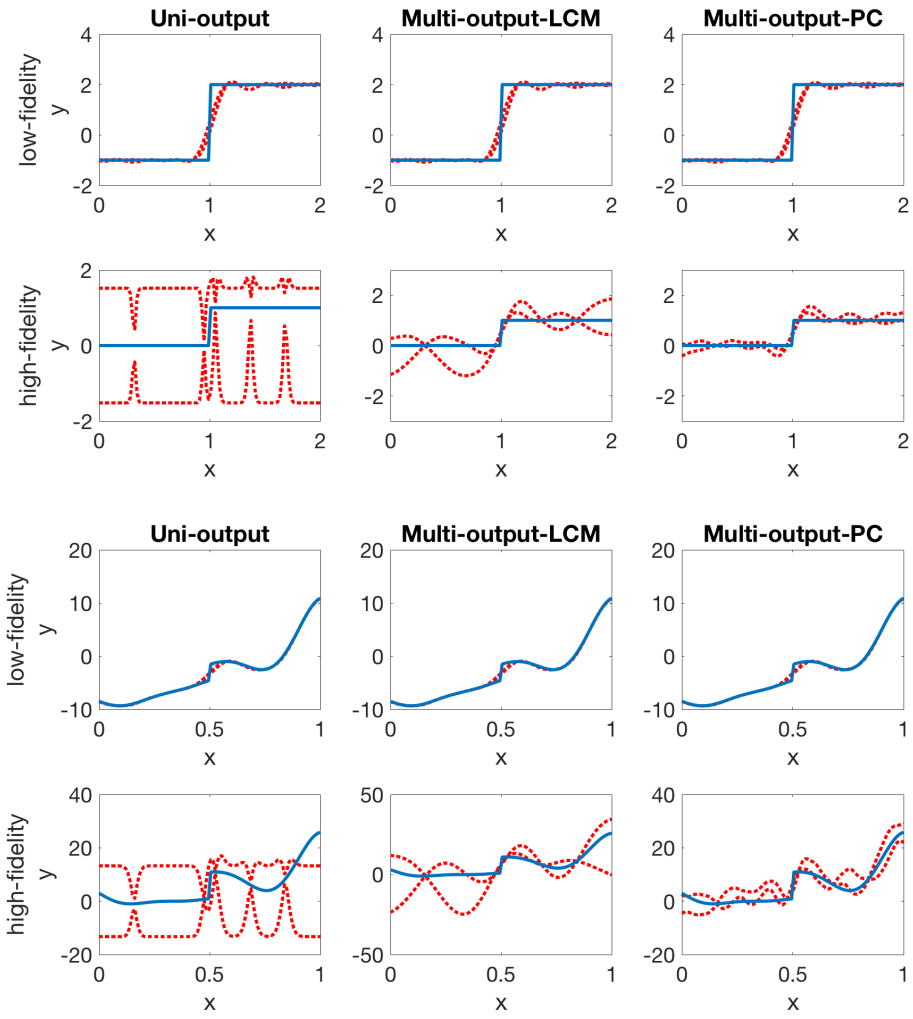

Figure 6. Multi-fidelity modelling: Plot of the Step function (top) and Forrester function (bottom) in blue. The Gaussian Process 95% credible intervals are given by dotted red lines in each case.

respectively. Ideally, we would prefer to take ϵ as small as possible but this complicates the numerical approximation of the solution.

The functions considered and corresponding posteriors are given in Fig. 7 and 8, while the uni-output and multi-output BQ estimates for integration of these functions against a uniform measure Π are given in the table in Fig. 9. Integer points between 0 and 10 are evaluated, with points at 2, 5 and 8 being used to evaluate the high fidelity model and the others used for the low fidelity model. The choice of kernel hyperparameters is made by maximising the marginal likelihood. Clearly, the two multi-output BQ algorithm give posteriors on the high-fidelity model which are much more concentrated on the true function than the uni-output BQ.

D.3. Global illumination problem

In Fig. 10, we plot the evolution of the worst-case integration error as N increases for the uni-output, two-output and five-output BQ with LMC kernel. As expected from Thm 1, the convergence occurs at the same rate in N but with a smaller rate constant the more outputs there are.

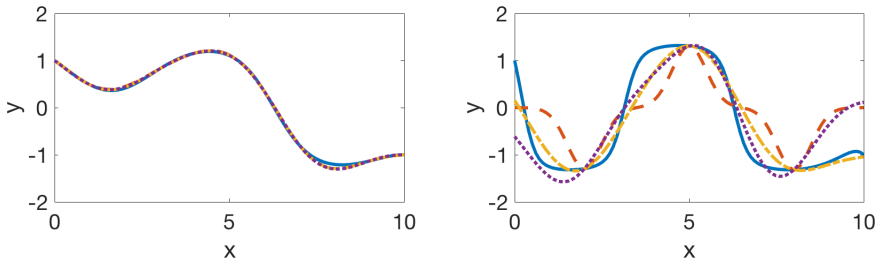

Figure 7. Multi-fidelity modelling: Plot of the solutions of Equation 4 for $\epsilon = 2$ (left) and $\epsilon = 0.1$ (right). Each plot gives the true function (blue) and their uni-output (dashed, red), LMC-based multi-output (dashed, yellow) and PC-based multi-output (dotted purple) approximations.

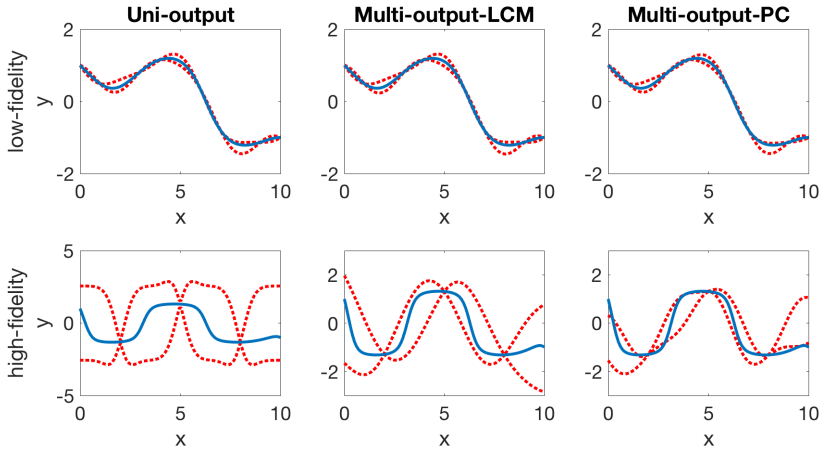

Figure 8. Multi-fidelity modelling: Plot of the solutions of Equation 4 for $\epsilon = 2$ (top) and $\epsilon = 0.1$ (bottom) in blue. The Gaussian Process 95% credible intervals are given by dotted red lines in each case.

Model	BQ	LMC-BQ	PC-BQ
AC (l)	0.004 (0.197)	0.006 (0.187)	0.007 (0.388)
AC (h)	0.211 (0.27)	0.002 (0.444)	0.037 (0.191)

Figure 9. Multi-fidelity modelling: Performance of uni-output BQ and multi-output BQ (with LMC and PC kernels) on the Allen-Cahn problem (AC) both for the low fidelity (l) and high fidelity (h) cases.

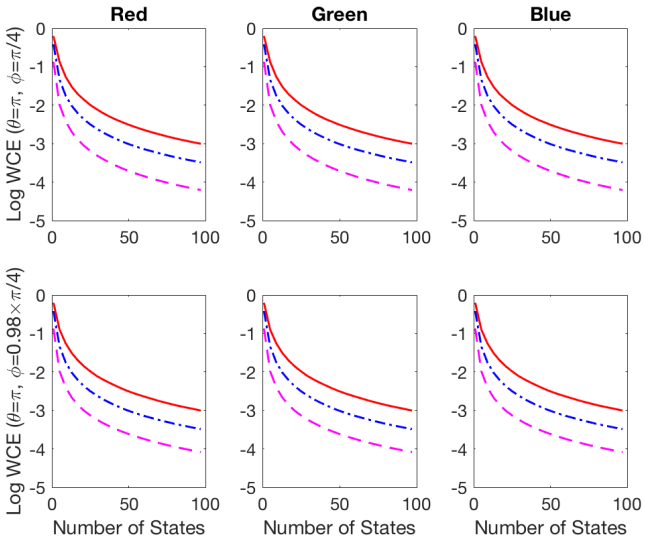

Figure 10. Global illumination: Plot of the worst-case integration error for f_1, f_2 in the case of the red, blue and green channels. Uni-output BQ is given in red (full line) whilst two-output BQ based on LMC is given in blue (dashed and dotted line) and five-output BQ based on LMC is given in magenta (dashed line).