The Coefficient of Determination of a Reduced Order Model

Robert A. Milton, Solomon F. Brown, Aaron S. Yeardley

Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

Abstract

Keywords: Global Sensitivity Analysis, Sobol' Index, Surrogate Model, Active Subspace

1. Introduction

2. Coefficient of Determination

Given a regression model

$$\text{Integrable } y\colon [0,1]^{M+1} \mapsto \mathbb{R}^L$$

begin by taking a uniformly distributed random variable (RV)

$$\mathsf{u} \sim \mathsf{U}\big([0]_{\mathbf{M+1}}, [1]_{\mathbf{M+1}}\big) := \mathsf{U}(0,1)^{M+1}$$

Exponentiation is categorical – repeated cartesian $\times$ or tensor $\otimes$ – unless otherwise specified. Square bracketed quantities are tensors, carrying their dimensions as a von Neumann ordinal subscript, in this case

$$\mathbf{M+1} := (0,\ldots,M) \supseteq \mathbf{m+1} := (0,\ldots,m \le M)$$

with void $\mathbf{0}$ voiding any tensor it subscripts. Ordinals are concatenated by Cartesian $\times$ and may be subtracted like sets, as in $\mathbf{M} - \mathbf{m} := (m,\ldots,M-1)$.

Email addresses: `r.a.milton@sheffield.ac.uk` (Robert A. Milton), `s.f.brown@sheffield.ac.uk` (Solomon F. Brown), `asyeardley1@sheffield.ac.uk` (Aaron S. Yeardley)

Subscripts refer to the tensor prior to any superscript operation, so $[y(\mathsf{u})]^2_{\mathbf{L}}$ is an $\mathbf{L}^2 := \mathbf{L} \times \mathbf{L}$ tensor, for example. The preference throughout this work is for uppercase constants and lowercase variables, in case of ordinals the lowercase ranging over the uppercase. We prefer o for an unbounded positive integer, to avoid O.

Expectations and variances will be subscripted by the dimensions of u marginalized (integrated over). Conditioning on the remaining dimensions is left implicit after Eq. (1), to lighten notation. Now, Construct $M + 1$ stochastic processes (SPs)

$$[\mathsf{y_m}]_{\mathbf{L}} := \mathbb{E}_{\mathbf{M-m}}[y(\mathsf{u})] := \mathbb{E}_{\mathbf{M-m}}\big[y(\mathsf{u})\,\big|\,[u]_{\mathbf{m}}\big] \tag{1}$$

ranging from $[\mathsf{y_0}]_{\mathbf{L}}$ to $[\mathsf{y_M}]_{\mathbf{L}}$. Every SP quietly depends on the ungovernable noise dimension $[\mathsf{u}]_M \perp [\mathsf{u}]_{\mathbf{M}}$, while ignoring governed input dimensions $[\mathsf{u}]_{\mathbf{M-m}}$. Sans serif symbols such as u, y generally refer to RVs and SPs, italic u, y being reserved for (tensor) functions and variables.

Following Daniell-Kolmogorov [1] pp.124 we may regard an SP as a random function, from which we shall freely extract finite dimensional distributions generated by a design matrix $[u]_{\mathbf{M \times o}}$ of $o \in \mathbb{Z}^+$ samples. Daniell-Kolmogorov incidentally secures u. Because y is (Lebesgue) integrable it must be measurable, guaranteeing $[\mathsf{y_0}]_{\mathbf{L}}$. Because all probability measures are finite, integrability of y implies integrability of y^n for all $n \in \mathbb{Z}^+$ [2]. So Fubini's Theorem [3] pp.77 allows all expectations to be taken in any order. This is sufficient to ensure every object appearing in this work.

Our aim is to compare predictions from a reduced model $\mathsf{y_m}$ with those of the full model $\mathsf{y_M}$. Correlation between these predictions is squared – using element-wise (Hadamard) multiplication $\circ$ and division – to form an RV called the coefficient of determination

$$\big[\mathsf{R^2_m}\big]_{\mathbf{L}^2} := \frac{\mathbb{V}_{\mathbf{M}}[\mathsf{y_m}, \mathsf{y_M}] \circ \mathbb{V}_{\mathbf{M}}[\mathsf{y_m}, \mathsf{y_M}]}{\mathbb{V}_{\mathbf{m}}[\mathsf{y_m}] \circ \mathbb{V}_{\mathbf{M}}[\mathsf{y_M}]} = \frac{\mathbb{V}_{\mathbf{m}}[\mathsf{y_m}]}{\mathbb{V}_{\mathbf{M}}[\mathsf{y_M}]} =: [\mathsf{S_m}]_{\mathbf{L}^2} \tag{2}$$

The closed Sobol' index is the complement of the commonplace total Sobol' index

$$[\mathsf{S_m}]_{\mathbf{L}^2} =: [1]_{\mathbf{L}^2} - \big[\mathsf{S^T_{M-m}}\big]_{\mathbf{L}^2}$$

It has mean value over the ungovernable noise dimension of

$$[S_{\mathbf{m}}]_{\mathbf{L}^2} := \mathbb{E}_M[\mathsf{S_m}] = \frac{V_{\mathbf{m}}}{V_{\mathbf{M}}} \tag{3}$$

$$\text{where } [V_{\mathbf{m}}]_{\mathbf{L}^2} := \mathbb{E}_M \mathbb{V}_{\mathbf{m}}[\mathsf{y_m}] \quad \forall \mathbf{m} \subseteq \mathbf{M} \tag{4}$$

and variance due to ungovernable noise of

$$[T_{\mathbf{m}}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathsf{S}_{\mathbf{m}}] = \frac{V_{\mathbf{m}}^2}{V_{\mathbf{M}}^2} \circ \left(\frac{W_{\mathbf{mm}}}{V_{\mathbf{m}}^2} - 2\frac{W_{\mathbf{mM}}}{V_{\mathbf{m}} \otimes V_{\mathbf{M}}} + \frac{W_{\mathbf{MM}}}{V_{\mathbf{M}}^2} \right) \quad (5)$$

$$\text{where} \quad [W_{\mathbf{mm}'}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathbf{m}}], \mathbb{V}_{\mathbf{m}'}[\mathsf{y}_{\mathbf{m}'}]] \quad \forall \mathbf{m}, \mathbf{m}' \subseteq \mathbf{M} \quad (6)$$

The remainder of this paper is devoted to calculating these two quantities – the coefficient of determination and its variance over noise (measurement error, squared).

3. Stochastic Process Estimates

Define noiseless mean predictors (functions of $[\mathsf{u}]_{\mathbf{m}}$) and their error SPs as

$$\begin{aligned}[f_{\mathbf{m}}]_{\mathbf{L}} &:= \mathbb{E}_{\mathbf{M}+\mathbf{1}-\mathbf{m}}[\mathsf{y}_{\mathbf{m}}]_{\mathbf{L}} \\ [e_{\mathbf{m}}]_{\mathbf{L}} &:= [\mathsf{y}_{\mathbf{m}}]_{\mathbf{L}} - [f_{\mathbf{m}}]_{\mathbf{L}} \end{aligned} \quad (7)$$

Adopt as design matrix a triad of inputs, their mutual dependence a function of conditioning, labelled according to how they will be marginalized

$$[\mathsf{u}]_{\mathbf{M}+\mathbf{1}\times\mathbf{3}} := [\mathsf{u}_0, \mathsf{u}_{\mathbf{m}}, \mathsf{u}_{\mathbf{M}}]$$

The triad elicits the response

$$[e]_{\mathbf{L}\times\mathbf{3}} := y([\mathsf{u}]_{\mathbf{M}+\mathbf{1}\times\mathbf{3}}) - \mathbb{E}_{\mathbf{M}+\mathbf{1}} \mathbb{E}_{\mathbf{M}'+\mathbf{1}'-\mathbf{m}'} \mathbb{E}_{M''}[y([\mathsf{u}]_{\mathbf{M}+\mathbf{1}\times\mathbf{3}})] \quad (8)$$

Primes mark independent input axes, otherwise each expectation applies to all three members of $[\mathsf{u}]_{\mathbf{M}+\mathbf{1}}^3$. It is not always obvious whether axes are independent or shared by the triad, but this can be mechanically checked against the measure of integration behind an expectation. Repeated expectations over the same axis are rare here, usually indicating that apparent repetitions must be "primed".

This purpose of the triad is to interrogate its response for moments in respect of shared noise

$$\begin{aligned}[\mu_n]_{(\mathbf{L}\times\mathbf{3})^n} &:= \mathbb{E}_M\big[[e]_{\mathbf{L}\times\mathbf{3}}^n\big] \\ [\mu_1]_{\mathbf{L}\times\mathbf{3}} &= [0]_{\mathbf{L}\times\mathbf{3}} \end{aligned} \quad (9)$$

for these embody

$$[\mu_{\mathbf{m}'...\mathbf{m}^{n'}}]_{\mathbf{L}^n} := [\mu_n]_{(\mathbf{L}\times i)^n} = \mathbb{E}_M[[e_{\mathbf{m}'}]_{\mathbf{L}} \otimes \cdots \otimes [e_{\mathbf{m}^{n'}}]_{\mathbf{L}}]$$

3

where $\mathbf{m}^{i\prime} \in \{\mathbf{0}, \mathbf{m}, \mathbf{M}\}$. This expression, along with $f_\mathbf{m}$, underpins both quantities we seek. The reduction which follows repeatedly realizes

$$[\mu_{\mathbf{0}...\mathbf{0}\mathbf{m}^{j\prime}...\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := \mathbb{E}_\mathbf{M}[\mu_{\mathbf{M}...\mathbf{M}\mathbf{m}^{j\prime}...\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} = \mathbb{E}_\mathbf{m}[\mu_{\mathbf{m}...\mathbf{m}\mathbf{m}^{j\prime}...\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} \qquad (10)$$

The expected conditional variance in Eq. (3) amounts to

$$\begin{aligned}
[V_\mathbf{m}]_\mathbf{L} &= \mathbb{E}_\mathbf{m}\mathbb{E}_M\big[[\mathsf{e}_\mathbf{m} + f_\mathbf{m}]_\mathbf{L}^2\big] - \mathbb{E}_M\big[[\mathsf{e}_\mathbf{0} + f_\mathbf{0}]_\mathbf{L}^2\big] \\
&= \mathbb{E}_\mathbf{m}\big[[f_\mathbf{m}]_\mathbf{L}^2\big] - [f_\mathbf{0}]_\mathbf{L}^2 + \mathbb{E}_\mathbf{m}[\mu_{\mathbf{mm}}]_{\mathbf{L}^2} - [\mu_{\mathbf{00}}]_{\mathbf{L}^2} \qquad (11) \\
&= \mathbb{E}_\mathbf{m}\big[[f_\mathbf{m}]_\mathbf{L}^2\big] - [f_\mathbf{0}]_\mathbf{L}^2
\end{aligned}$$

The covariance between conditional variances in Eq. (5) is

$$\begin{aligned}
[W_{\mathbf{mm}\prime}]_{\mathbf{L}^4} &:= \mathbb{V}_M[\mathbb{V}_\mathbf{m}[\mathsf{y}_\mathbf{m}], \mathbb{V}_{\mathbf{m}\prime}[\mathsf{y}_{\mathbf{m}\prime}]] \\
&= \mathbb{V}_M\big[\mathbb{E}_\mathbf{m}\big[[\mathsf{y}_\mathbf{m}]_\mathbf{L}^2 - [\mathsf{y}_\mathbf{0}]_\mathbf{L}^2\big], \mathbb{E}_{\mathbf{m}\prime}\big[[\mathsf{y}_{\mathbf{m}\prime}]_\mathbf{L}^2 - [\mathsf{y}_\mathbf{0}]_\mathbf{L}^2\big]\big] \\
&= \mathbb{E}_M\big[\mathbb{E}_\mathbf{m}\big[[\mathsf{y}_\mathbf{m}]_\mathbf{L}^2 - [\mathsf{y}_\mathbf{0}]_\mathbf{L}^2\big] \otimes \mathbb{E}_{\mathbf{m}\prime}\big[[\mathsf{y}_{\mathbf{m}\prime}]_\mathbf{L}^2 - [\mathsf{y}_\mathbf{0}]_\mathbf{L}^2\big]\big] \\
&\quad - [V_\mathbf{m}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}\prime}]_{\mathbf{L}^2} \\
&= [A_{\mathbf{mm}\prime} - A_{\mathbf{0m}\prime} - A_{\mathbf{m0}} + A_{\mathbf{00}}]_{\mathbf{L}^4}
\end{aligned}$$

Here, for any $\mathbf{m}, \mathbf{m}\prime \subseteq \mathbf{M}$

$$\begin{aligned}
[A_{\mathbf{mm}\prime}]_{\mathbf{L}^4} &:= \mathbb{E}_\mathbf{m}\mathbb{E}_{\mathbf{m}\prime}\mathbb{E}_M\big[[\mathsf{y}_\mathbf{m}]_\mathbf{L}^2 \otimes [\mathsf{y}_{\mathbf{m}\prime}]_\mathbf{L}^2\big] - [V_\mathbf{m}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}\prime}]_{\mathbf{L}^2} \\
&= \mathbb{E}_\mathbf{m}\mathbb{E}_{\mathbf{m}\prime}\mathbb{E}_M\big[[\mathsf{e}_\mathbf{m} + f_\mathbf{m}]_\mathbf{L}^2 \otimes [\mathsf{e}_{\mathbf{m}\prime} + f_{\mathbf{m}\prime}]_\mathbf{L}^2\big] - [V_\mathbf{m}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}\prime}]_{\mathbf{L}^2}
\end{aligned}$$

exploiting the fact that $V_\mathbf{0} = [0]_{\mathbf{L}^2}$. Equation (10) cancels all terms beginning with $[\mathsf{e}_\mathbf{m}]_\mathbf{L}^2$, first across $A_{\mathbf{m},\mathbf{m}\prime} - A_{\mathbf{0},\mathbf{m}\prime}$ then across $A_{\mathbf{m},\mathbf{0}} - A_{\mathbf{0},\mathbf{0}}$. All remaining terms ending in $[f_{\mathbf{m}\prime}]_\mathbf{L}^2$ are eliminated by error centralization $\mu_1 = 0$ and

$$\begin{aligned}
\mathbb{E}_\mathbf{m}\mathbb{E}_{\mathbf{m}\prime}\big[[f_\mathbf{m}]_\mathbf{L}^2 \otimes [f_{\mathbf{m}\prime}]_\mathbf{L}^2\big] - [V_\mathbf{m}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}\prime}]_{\mathbf{L}^2} \\
= [V_\mathbf{m}]_{\mathbf{L}^2} \otimes [f_\mathbf{0}]_\mathbf{L}^2 + [f_\mathbf{0}]_\mathbf{L}^2 \otimes [V_{\mathbf{m}\prime}]_{\mathbf{L}^2} + [f_\mathbf{0}]_\mathbf{L}^4
\end{aligned}$$

cancelling across $A_{\mathbf{m},\mathbf{m}\prime} - A_{\mathbf{0},\mathbf{m}\prime} - A_{\mathbf{m},\mathbf{0}} + A_{\mathbf{0},\mathbf{0}}$. Similar arguments eliminate $[\mathsf{e}_{\mathbf{m}\prime}]_\mathbf{L}^2$ and $[f_\mathbf{m}]_\mathbf{L}^2$. Effectively then

$$[A_{\mathbf{mm}\prime}]_{\mathbf{L}^4} = \sum_{\pi(\mathbf{L}^2)} \sum_{\pi(\mathbf{L}\prime^2)} \mathbb{E}_\mathbf{m}\mathbb{E}_{\mathbf{m}\prime}[f_\mathbf{m} \otimes \mu_{\mathbf{mm}\prime} \otimes f_{\mathbf{m}\prime}]_{\mathbf{L}^2 \times \mathbf{L}\prime^2} \qquad (12)$$

where each summation is over permutations of tensor axes

$$\pi(\mathbf{L}^2) := \{(\mathbf{L} \times \mathbf{L}\prime\prime), (\mathbf{L}\prime\prime \times \mathbf{L})\} \quad ; \quad \pi(\mathbf{L}\prime^2) := \{(\mathbf{L}\prime \times \mathbf{L}\prime\prime\prime), (\mathbf{L}\prime\prime\prime \times \mathbf{L}\prime)\}$$

Primes on constants are for bookeeping purposes only ($\mathbf{L}^{i\prime} = \mathbf{L}$ always), they do not change the value of the constant – unlike primes on variables ($\mathbf{m}^{i\prime}$ need not equal $\mathbf{m}$ in general). In order to further elucidate these estimates, we must fill in the details of the underlying stochastic processes, sufficiently identifying the regression y by its first two moments f, μ_2.

4. Interlude: Gaussian Process (GP) Regression

A GP over x is formally defined and specified by

$$[\mathsf{y}_\mathsf{M}]_\mathbf{L} \mid [x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\left([\bar{y}(x)]_{\mathbf{L}\times\mathbf{o}}, [k_\mathsf{y}(x,x)]_{(\mathbf{L}\times\mathbf{o})^2}\right) \quad \forall o \in \mathbb{Z}^+$$

where tensor ranks concatenate into a multivariate normal distribution

$$[]_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}^\dagger\left([]_{\mathbf{L}\times\mathbf{o}}, []_{(\mathbf{L}\times\mathbf{o})^2}\right) \iff []^\dagger_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\left([]^\dagger_{\mathbf{L}\times\mathbf{o}}, []^\dagger_{(\mathbf{L}\times\mathbf{o})^2}\right)$$

$$\left[[]^\dagger_{\mathbf{L}\times\mathbf{o}}\right]_{\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o}} := []_{l-1\times\mathbf{o}}$$

$$\left[[]^\dagger_{(\mathbf{L}\times\mathbf{o})^2}\right]_{(\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o})\times(\mathbf{l'o}-(\mathbf{l'}-\mathbf{1})\mathbf{o})} := []_{l-1\times\mathbf{o}\times l'-1\times\mathbf{o}}$$

supporting the fundamental definition of the Gaussian process kernel, as a covariance over response space

$$[k_\mathsf{y}(x,x)]_{l\times o\times l'\times o'} := \mathbb{V}_{\mathbf{Lo}}\left[[\mathsf{y}_\mathsf{M}|x]_{l\times o}, [\mathsf{y}_\mathsf{M}|x]_{l'\times o'}\right]$$

4.1. Tensor Gaussians

A tensor Gaussian like $\mathrm{p}\left([x]_{\mathbf{m}\times\mathbf{o}}\mid [x']_{\mathbf{m}\times\mathbf{o'}}, [\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\right)$ is defined element-wise

$$\left[\mathrm{p}\left([x]_{\mathbf{m}\times\mathbf{o}}\mid [x']_{\mathbf{m}\times\mathbf{o'}}, [\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\right)\right]_{l\times o\times l'\times o'} := (2\pi)^{-M/2}\left|[\Sigma]_{l\times l'}\right|^{-1/2}$$

$$\exp\left(-\frac{[x-x']^\mathsf{T}_{\mathbf{m}\times l\times o\times l'\times o'}[\Sigma]^{-1}_{l\times l'}[x-x']_{\mathbf{m}\times l\times o\times l'\times o'}}{2}\right) \quad (13)$$

in terms of the matrix

$$[\Sigma]_{l\times l'} := [\Sigma]_{l\times l'\times\mathbf{m}^2}$$

and the transpose $^\mathsf{T}$ (moving first multi-index to last) of the broadcast difference between two tensors

$$[x-x']_{\mathbf{m}\times l\times o\times l'\times o'} := [x]_{\mathbf{m}\times o} - [x']_{\mathbf{m}\times o'}$$

5

Remarkably, the algebraic development in the remainder of this paper relies almost exclusively on an invaluable product formula reported in [4]:

$$\mathrm{p}(z\,|\,a, A) \circ \mathrm{p}(\Theta^\mathsf{T} z\,|\,b, B) = \mathrm{p}(0\,|\,(b - \Theta^\mathsf{T} a), (B + \Theta^\mathsf{T} A\Theta))$$
$$\circ\, \mathrm{p}\big(z\,|\,(A^{-1} + \Theta B^{-1}\Theta^\mathsf{T})^{-1}(A^{-1}a + \Theta B^{-1}b), (A^{-1} + \Theta B^{-1}\Theta^\mathsf{T})^{-1}\big) \quad (14)$$

This formula and the Gaussian tensors behind it will appear in a variety of guises.

4.2. Prior GP

GP regression decomposes output $[\mathsf{y_M}]_\mathbf{L}$ into signal GP $[\mathsf{f_M}]_\mathbf{L}$, and independent noise GP $[\mathsf{e}_M]_\mathbf{L}$ with constant noise covariance $[E]_{\mathbf{L}^2}$

$$[\mathsf{y_M}|E]_\mathbf{L} = [\mathsf{f_M}]_\mathbf{L} + [\mathsf{e}_M|E]_\mathbf{L}$$
$$[\mathsf{e}_M|E]_\mathbf{L}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}}, [E]_{(\mathbf{L}\times 1)^2} \circ \langle 1\rangle_{(1\times\mathbf{o})^2}\Big)$$

The RBF kernel is hyperparametrized by signal covariance $[F]_{\mathbf{L}^2}$ and the matrix $[\Lambda]_{\mathbf{L}\times\mathbf{M}}$ of characteristic lengthscales for each output/input combination. Angle brackets denoting a (perhaps broadcast) diagonal tensor, such as the identity matrix $\langle 1\rangle_{\mathbf{m}^2} =: \langle [1]_\mathbf{m}\rangle$, are used to define

$$\big\langle \Lambda^2 \pm I\big\rangle_{l\times l'\times\mathbf{M}^2} := \big\langle [\Lambda]_{l\times\mathbf{M}} \circ [\Lambda]_{l'\times\mathbf{M}} \pm [I]_\mathbf{M}\big\rangle \qquad I \in \{0\} \cup \mathbb{Z}^+$$
$$\big\langle \Lambda^2\big\rangle_{l\times l'\times\mathbf{M}^2} := \big\langle \Lambda^2 \pm 0\big\rangle_{l\times l'}$$
$$[\pm F]_{l\times l'} := (2\pi)^{M/2}\left|\big\langle\Lambda^2\big\rangle_{l\times l'}\right|^{1/2} [F]_{l\times l'}$$

and implement the objective RBF prior using Eq. (13)

$$[\mathsf{f_M}|F, \Lambda]_\mathbf{L}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}}, [\pm F]_{(\mathbf{L}\times 1)^2} \circ \mathrm{p}\big([x]_{\mathbf{M}\times\mathbf{o}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}, \big\langle\Lambda^2\big\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\big)\Big)$$

4.3. Predictive GP

Bayesian inference for GP regression further conditions the hyper-parametrized GP $\mathsf{y}|E, F, \Lambda$ on the observed realization of the random variable $[\mathsf{y}|X]$

$$[Y]^\dagger_{\mathbf{L}\times\mathbf{N}} := \big[[\mathsf{y_M}|E, F, \Lambda]_\mathbf{L}\,\big|\,[X]_{\mathbf{M}\times\mathbf{N}}\big]^\dagger(\omega) \in \mathbb{R}^{LN}$$

To this end we define

$$[K_e]_{\mathbf{Lo} \times \mathbf{Lo}} := \mathbb{V}_{\mathbf{Lo}} \Big[\big[[e_M | E]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M} \times \mathbf{o}} \big]^\dagger \Big]$$

$$= \Big[[E]_{(\mathbf{L} \times 1)^2} \circ \langle 1 \rangle_{(1 \times \mathbf{o})^2} \Big]^\dagger$$

$$[k(x, x')]_{\mathbf{Lo} \times \mathbf{L'o'}} := \mathbb{V}_{\mathbf{Lo}} \Big[\big[[f_M | F, \Lambda]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M} \times \mathbf{o}} \big]^\dagger, \big[[f_M | F, \Lambda]_{\mathbf{L'}} \,\big|\, [x']_{\mathbf{M} \times \mathbf{o'}} \big]^\dagger \Big] \quad (15)$$

$$= \Big[[\pm F]_{\mathbf{L}^2} \circ \mathrm{p} \Big([x]_{\mathbf{M} \times \mathbf{o}} \big| [x']_{\mathbf{M} \times \mathbf{o'}}, \langle \Lambda^2 \rangle_{\mathbf{L}^2 \times \mathbf{M}^2} \Big) \Big]^\dagger$$

$$[K_Y]_{\mathbf{LN} \times \mathbf{LN}} := \mathbb{V}_{\mathbf{Lo}} \Big[\big[[y | E, F, \Lambda]_{\mathbf{L}} \,\big|\, [X]_{\mathbf{M} \times \mathbf{N}} \big]^\dagger \Big]$$

$$= k([X]_{\mathbf{M} \times \mathbf{N}}, [X]_{\mathbf{M} \times \mathbf{N}}) + [K_e]_{\mathbf{LN} \times \mathbf{LN}}$$

Applying Bayes' rule

$$\mathrm{p}(f_M | Y) \mathrm{p}(Y) = \mathrm{p}(Y | f_M) \mathrm{p}(f_M) = \mathrm{p}\big(Y^\dagger \big| f_M{}^\dagger, K_{e_M} \big) \, \mathrm{p}\big(f_M{}^\dagger \big| [0]_{\mathbf{LN}}, k(X, X) \big)$$

$$= \mathrm{p}\big(f_M{}^\dagger \big| Y^\dagger, K_{e_M} \big) \, \mathrm{p}\big(f_M{}^\dagger \big| [0]_{\mathbf{LN}}, k(X, X) \big)$$

Eq. (14) immediately reveals the marginal likelihood

$$\mathrm{p}\big([Y | E, F, \Lambda] \,\big|\, X \big) = \mathrm{p}\Big([Y]_{\mathbf{L} \times \mathbf{N}}^\dagger \,\Big|\, [0]_{\mathbf{LN}}, K_Y \Big) \quad (16)$$

and the posterior distribution

$$\big[[f_M | Y | E, F, \Lambda] \,\big|\, X \big]_{\mathbf{L} \times \mathbf{N}}^\dagger \sim$$
$$\mathsf{N}\big(k(X, X) K_Y^{-1} Y^\dagger, \; k(X, X) - k(X, X) K_Y^{-1} k(X, X) \big)$$

The ultimate goal is the posterior predictive GP which extends the posterior distribution to arbitrary – usually unobserved – $[x]_{\mathbf{M} \times \mathbf{o}}$. This is traditionally derived from the definition of conditional probability, but this seems unnecessary, for the extension must recover the posterior distribution when $x = X$. There is only one way of selectively replacing X with x in the posterior formula which preserves the coherence of tensor ranks:

$$\big[[f_M | Y | E, F, \Lambda] \,\big|\, x \big]_{\mathbf{L} \times \mathbf{o}}^\dagger \sim$$
$$\mathsf{N}\big(k(x, X) K_Y^{-1} Y^\dagger, \; k(x, x) - k(x, X) K_Y^{-1} k(X, x) \big) \quad (17)$$

4.4. GP Optimization

Henceforth we implicitly condition on optimal hyperparameters, which maximise the marginal likelihood Eq. (16).

$$[E]_{\mathbf{L^2}}, [F]_{\mathbf{L^2}}, [\Lambda]_{\mathbf{L \times M}} := \operatorname{argmax} \operatorname{p}\left([Y]_{\mathbf{L \times N}}^{\dagger} \,\middle|\, [0]_{\mathbf{LN}}, K_Y \right) \qquad (18)$$

The lengthscale tensor could feasibly have been of maximal rank $[\Lambda]_{\mathbf{L^2 \times M}}$. We have restricted this to $[\Lambda]_{\mathbf{L \times M}}$, as one set of ARD lengthscales per output is heuristically satisfying and enables effective optimization as follows. For each output $l \in \mathbf{L}$ construct a separate GP to optimize the diagonal hyperparameters

$$[E]_{l^2}, [F]_{l^2}, [\Lambda]_{l \times \mathbf{M}} = \operatorname{argmax} \operatorname{p}\left([Y]_{l \times \mathbf{N}}^{\dagger} \,\middle|\, [0]_{\mathbf{N}}, [K_Y]_{(l+1)\mathbf{N}-\mathbf{N} \times (l+1)\mathbf{N}-\mathbf{N}} \right)$$

From this starting point, E, F may be optimized (off-diagonal elements in particular) in the full multi-output GP Eq. (18). One may then attempt to re-optimize lengthscales according to Eq. (18), and iterate, although this may be gilding the lily.

5. Gaussian Process Estimates

This Section fills out the stochastic process estimates for the coefficient of determination in the case of Gaussian Process Regression. Let $c \colon \mathbb{R} \to [0, 1]$ be the (bijective) CDF of the standard, univariate normal distribution, and define the triads

$$[\mathsf{z}]_{\mathbf{M \times 3}} := c^{-1}\left([\mathsf{u}]_{\mathbf{M \times 3}} \right) \sim \mathsf{N}\left([0]_{\mathbf{M+1 \times 3}}, \langle 1 \rangle_{(\mathbf{M+1})^2} \right)$$

$$[\mathsf{x}]_{\mathbf{M' \times 3}} := [\Theta]_{\mathbf{M \times M'}}^{\mathsf{T}} [\mathsf{z}]_{\mathbf{M \times 3}}$$

Here, the rotation matrix $[\Theta]_{\mathbf{M \times M'}}^{\mathsf{T}} = [\Theta]_{\mathbf{M \times M'}}^{-1}$ is broadcast to multiply the triad $[\mathsf{z}]_{\mathbf{M \times 3}}$. Throughout the remainder of this paper, primed ordinal subscripts are used to specify einstein summed (contracted) products of tensors. The purpose of ths arbitrary rotation is to allow GPs whose input basis x is not aligned with the fundamental basis u of model reduction and coefficient of determination. The latter is aligned with z which is the input we must condition.

Adding shared gaussian noise $[\mathsf{e}_M | E]_{\mathbf{L}}$ to Eq. (17) yields

$$\left[y([\mathsf{u}]_{\mathbf{M+1\times3}})\,\big|\,[u]_{\mathbf{M\times3}}\right]^{\dagger}_{\mathbf{L\times3}} = \left[[\mathsf{y_M}|Y|E,F,\Lambda]\,\big|[z]_{\mathbf{M\times3}}\right]^{\dagger}_{\mathbf{L\times3}} \sim$$
$$\mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x) - k(x,X)K_Y^{-1}k(X,x) + E^{\dagger}\big) \quad (19)$$

using broadcast $[E^{\dagger}]_{\mathbf{L3\times L3}} := [[E]_{(\mathbf{L}\times1)^2} \circ [1]_{(1\times\mathbf{3})^2}]^{\dagger}_{(\mathbf{L\times3})^2}$. From the development in Section 3, especially Eqs. (7) to (9), the triad response has mean f and variance $\mu2$ and this is all the information we need to compute the coefficient of determination and its variance. Obviously these can be read straight off Eq. (19).

For the sake of completeness, and elucidation we bring the GP estimate fully under the umbrella of the SP estimate by identifying ungovernable noise, ultimately ascribable to $[\mathsf{u}]_M$. let $d_{\mathbf{L}}\colon (0,1) \to (0,1)^L$ concatenate every L^{th} decimal place starting at l, for each dimension $l \le L$ of $(0,1)^L$, then Eq. (19) can be rewritten

$$y([\mathsf{u}]_{\mathbf{M+1\times3}})\,\big|\,[u]_{\mathbf{M\times3}} = [f]_{\mathbf{L\times3}} + [\mu_2]^{/\mathsf{T}}_{(\mathbf{L\times3})^2}\left[c^{-1}(d_{\mathbf{L}}([\mathsf{u}]_M))\right]_{\mathbf{L\times3}} \quad (20)$$

6. Gaussian Process Estimates

This Section fills out the stochastic process estimates for the coefficient of determination in the case of Gaussian Process Regression. Let $c\colon \mathbb{R}^{Mo} \to [0,1]^{Mo}$ be the (bijective) CDF of the standard, univariate normal distribution, and let $d_{\mathbf{L}}\colon (0,1) \to (0,1)^L$ concatenate every L^{th} decimal place starting at l, for each dimension $l \le L$ of $(0,1)^L$. Then

$$[\mathsf{z}]_{\mathbf{M}} := c^{-1}([\mathsf{u}]_{\mathbf{M}}) \sim \mathsf{N}([0]_{\mathbf{M}}, \langle 1 \rangle_{\mathbf{M}^2})$$
$$[\mathsf{e}_M]_L := c^{-1}(d_{\mathbf{L}}([\mathsf{u}]_M)) \sim \mathsf{N}([0]_{\mathbf{L}}, \langle 1 \rangle_{\mathbf{L}^2})$$

Algebraic developments rely exclusively on trivial normal marginalization and scaling

$$[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}, [\Sigma]_{\mathbf{M\times M}}\big) \Rightarrow [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}, [\Sigma]_{\mathbf{m\times m}}\big)$$
$$[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}, [\Sigma]_{\mathbf{M\times M}}\big) \Rightarrow [\Theta]^{\mathsf{T}}_{\mathbf{M\times M}}[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}(\Theta^{\mathsf{T}}Z, \Theta^{\mathsf{T}}\Sigma\Theta) \quad (21)$$

$$(22)$$

$$[f_{\mathbf{m}}]_{\mathbf{L}} = \mathbb{E}_{\mathbf{M+1-m}}\big[k(\mathsf{x},X)K_Y^{-1}Y^{\dagger}\big|[\mathsf{z}]_{\mathbf{m}}\big] \quad (23)$$

and

$$[\mu_2]_{(\mathbf{L\times3})^2} =$$
$$\mathbb{E}_{\mathbf{M+1}}\mathbb{E}_{\mathbf{M+1-m}}\mathbb{E}_{M''}\big[k(\mathsf{x},\mathsf{x}) - k(\mathsf{x},X)K_Y^{-1}k(X,\mathsf{x}) + E^{\dagger}\big|[\mathsf{z}]_{\mathbf{M\times3}}\big] \quad (24)$$

9

References

[1] L. C. G. Rogers, D. Williams, Diffusions, Markov Processes, and Martingales, Cambridge University Press, 2000.

[2] A. Villani, Another note on the inclusion $L_p(\mu) \subset L_q(\mu)$, The American Mathematical Monthly 92 (7) (1985) 485. `doi:10.2307/2322503`.

[3] D. Williams, Probability with Martingales, Cambridge University Press, 1991.

[4] C. E. Rasmussen, Some useful gaussian and matrix equations (2016).
URL `http://mlg.eng.cam.ac.uk/teaching/4f13/1617/gaussian%20and%20matrix%20equations.pdf`