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

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

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

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 $[y(\mathsf{u})]_{\mathbf{L}}^2$ 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, avoiding O.

Expectations and variances will be subscripted by the dimensions of u marginalized. 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 $[u]_{\mathbf{M}\times\mathbf{o}}$ 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

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

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

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

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

Adopt as design matrix a triad of inputs to condition $[\mathsf{u}]_{\mathbf{M+1 \times 3}}$, eliciting the response

$$[\mathsf{y}]_{\mathbf{L \times 3}} := \mathbb{E}_{\mathbf{M}} \mathbb{E}_{\mathbf{M'-m'}} \mathbb{E}_{\mathbf{0''}} \left[y([\mathsf{u}]_{\mathbf{M+1 \times 3}}) \,\middle|\, [[u]_{\mathbf{0}}, [u]_{\mathbf{m'}}, [u]_{\mathbf{M''}}] \right] \quad (7)$$

Primes mark independent input axes, otherwise expectations are shared by all three members of the triad. 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". The purpose of the triad is to interrogate its response for moments in respect of ungovernable noise (which is shared)

$$[\mu_n]_{(\mathbf{L \times 3})^n} := \mathbb{E}_M\left[[\mathsf{y}]_{\mathbf{L \times 3}}^n\right] \quad \forall n \in \mathbb{Z}^+ \quad (8)$$

for these embody

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

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

$$[\mu_{\mathbf{0}\ldots\mathbf{0m}^{j'}\ldots\mathbf{m}^{n'}}]_{\mathbf{L}^n} := \mathbb{E}_{\mathbf{M}}[\mu_{\mathbf{M}\ldots\mathbf{Mm}^{j'}\ldots\mathbf{m}^{n'}}]_{\mathbf{L}^n} = \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}\ldots\mathbf{mm}^{j'}\ldots\mathbf{m}^{n'}}]_{\mathbf{L}^n} \quad (9)$$

Defining

$$[\mathsf{e}]_{\mathbf{L \times 3}} := \mathsf{y} - \mu_1 \quad (10)$$

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

$$[V_{\mathbf{m}}]_{\mathbf{L}} = \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_M\big[[\mathsf{e}_{\mathbf{m}} + \mu_{\mathbf{m}}]_{\mathbf{L}}^2\big] - \mathbb{E}_M\big[[\mathsf{e}_0 + \mu_0]_{\mathbf{L}}^2\big]$$
$$= \mathbb{E}_{\mathbf{m}}\big[[\mu_{\mathbf{m}}]_{\mathbf{L}}^2\big] - [\mu_0]_{\mathbf{L}}^2 + \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{mm}}]_{\mathbf{L}^2} - [\mu_{00}]_{\mathbf{L}^2} \tag{11}$$
$$= \mathbb{E}_{\mathbf{m}}\big[[\mu_{\mathbf{m}}]_{\mathbf{L}}^2\big] - [\mu_0]_{\mathbf{L}}^2$$

and the covariance between conditional variances in Eq. (5) is

$$[W_{\mathbf{mm'}}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathbf{m}}]\,, \mathbb{V}_{\mathbf{m'}}[\mathsf{y}_{\mathbf{m'}}]]$$
$$= \mathbb{V}_M\big[\mathbb{E}_{\mathbf{m}}\big[[\mathsf{y}_{\mathbf{m}}]_{\mathbf{L}}^2 - [\mathsf{y}_0]_{\mathbf{L}}^2\big]\,, \mathbb{E}_{\mathbf{m'}}\big[[\mathsf{y}_{\mathbf{m'}}]_{\mathbf{L}}^2 - [\mathsf{y}_0]_{\mathbf{L}}^2\big]\big]$$
$$= \mathbb{E}_M\big[\mathbb{E}_{\mathbf{m}}\big[[\mathsf{y}_{\mathbf{m}}]_{\mathbf{L}}^2 - [\mathsf{y}_0]_{\mathbf{L}}^2\big] \otimes \mathbb{E}_{\mathbf{m'}}\big[[\mathsf{y}_{\mathbf{m'}}]_{\mathbf{L}}^2 - [\mathsf{y}_0]_{\mathbf{L}}^2\big]\big] \tag{12}$$
$$\quad - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m'}}]_{\mathbf{L}^2}$$
$$= [A_{\mathbf{mm'}} - A_{\mathbf{0m'}} - A_{\mathbf{m0}} + A_{\mathbf{00}}]_{\mathbf{L}^4}$$

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

$$[A_{\mathbf{mm'}}]_{\mathbf{L}^4} := \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m'}}\,\mathbb{E}_M\big[[\mathsf{y}_{\mathbf{m}}]_{\mathbf{L}}^2 \otimes [\mathsf{y}_{\mathbf{m'}}]_{\mathbf{L}}^2\big] - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m'}}]_{\mathbf{L}^2}$$
$$= \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m'}}\,\mathbb{E}_M\big[[\mathsf{e}_{\mathbf{m}} + \mu_{\mathbf{m}}]_{\mathbf{L}}^2 \otimes [\mathsf{e}_{\mathbf{m'}} + \mu_{\mathbf{m'}}]_{\mathbf{L}}^2\big] - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m'}}]_{\mathbf{L}^2}$$

exploiting the fact that $V_{\mathbf{0}} = [0]_{\mathbf{L}^2}$. Equation (9) cancels all terms beginning with $[\mathsf{e}_{\mathbf{m}}]_{\mathbf{L}}^2$, first across $A_{\mathbf{m},\mathbf{m'}} - A_{\mathbf{0},\mathbf{m'}}$ then across $A_{\mathbf{m},\mathbf{0}} - A_{\mathbf{0},\mathbf{0}}$. All remaining terms ending in $[\mu_{\mathbf{m'}}]_{\mathbf{L}}^2$ are eliminated by error centralization $\mu_1 = 0$ and

$$\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m'}}\big[[\mu_{\mathbf{m}}]_{\mathbf{L}}^2 \otimes [\mu_{\mathbf{m'}}]_{\mathbf{L}}^2\big] - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m'}}]_{\mathbf{L}^2}$$
$$= [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [\mu_0]_{\mathbf{L}}^2 + [\mu_0]_{\mathbf{L}}^2 \otimes [V_{\mathbf{m'}}]_{\mathbf{L}^2} + [\mu_0]_{\mathbf{L}}^4$$

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

$$[A_{\mathbf{mm'}}]_{\mathbf{L}^4} = \sum_{\pi(\mathbf{L}^2)} \sum_{\pi(\mathbf{L'}^2)} \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m'}}[\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm'}} \otimes \mu_{\mathbf{m'}}]_{\mathbf{L}^2 \times \mathbf{L'}^2}$$
$$= 4\,\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m'}}[\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm'}} \otimes \mu_{\mathbf{m'}}]_{\mathbf{L}^4} \tag{13}$$

where each summation is over permutations of tensor axes

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

Primes on constants are for bookeeping purposes only ($\mathbf{L}^{i'} = \mathbf{L}$ always), they do not change the value of the constant – unlike primes on variables ($\mathbf{m}^{i'}$

4

need not equal $\mathbf{m}$ in general). In this case the four permutations all give the same result, by symmetry of covariances, hence the last equality

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 μ_1, μ_2.

4. Interlude: Gaussian Process (GP) Regression

A GP over x is formally defined and specified by

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

where tensor ranks concatenate into a multivariate normal distribution

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

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

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

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

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

4.1. Tensor Gaussians

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

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

$$\exp\left(-\frac{[x-x']^{\mathsf{T}}_{\mathbf{m} \times l \times o \times l' \times o'} [\Sigma]^{-1}_{l \times l'} [x-x']_{\mathbf{m} \times l \times o \times l' \times o'}}{2}\right) \quad (14)$$

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]:

$$p(z\,|\,a,A) \circ p(\Theta^\mathsf{T} z\,|\,b,B) = p(0\,|\,(b-\Theta^\mathsf{T} a),(B+\Theta^\mathsf{T} A\Theta))$$
$$\circ\, p\big(z\,|\,(A^{-1}+\Theta B^{-1}\Theta^\mathsf{T})^{-1}(A^{-1}a+\Theta B^{-1}b),(A^{-1}+\Theta B^{-1}\Theta^\mathsf{T})^{-1}\big) \quad (15)$$

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

4.2. Prior GP

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

$$[\mathsf{y}_\mathsf{M}|E]_\mathbf{L} = [\mathsf{f}_\mathsf{M}]_\mathbf{L} + [\mathsf{e}_M|E]_\mathbf{L}$$
$$[\mathsf{e}_M|E]_\mathbf{L}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}},[E]_{(\mathbf{L}\times1)^2}\circ\langle1\rangle_{(1\times\mathbf{o})^2}\Big)$$

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 $\langle1\rangle_{\mathbf{m}^2} =: \langle[1]_\mathbf{m}\rangle$, are used to define

$$\langle\Lambda^2\pm I\rangle_{l\times l'\times\mathbf{M}^2} := \langle[\Lambda]_{l\times\mathbf{M}}\circ[\Lambda]_{l'\times\mathbf{M}}\pm[I]_\mathbf{M}\rangle \qquad I\in\{0\}\cup\mathbb{Z}^+$$
$$\langle\Lambda^2\rangle_{l\times l'\times\mathbf{M}^2} := \langle\Lambda^2\pm0\rangle_{l\times l'}$$
$$[\pm F]_{l\times l'} := (2\pi)^{M/2}\left|\langle\Lambda^2\rangle_{l\times l'}\right|^{1/2}[F]_{l\times l'}$$

and implement the objective RBF prior using Eq. (14)

$$[\mathsf{f}_\mathsf{M}|F,\Lambda]_\mathbf{L}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}},[\pm F]_{(\mathbf{L}\times1)^2}\circ\mathrm{p}\big([x]_{\mathbf{M}\times\mathbf{o}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}},\langle\Lambda^2\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\big)\Big)$$

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 $[\mathsf{y}|X]$

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

To this end we define

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

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

$$[k(x, x')]_{\mathbf{Lo} \times \mathbf{L'o'}} := \mathbb{V}_{\mathbf{Lo}}\Big[\big[[f_M|F, \Lambda]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M \times o}}\big]^\dagger, \big[[f_M|F, \Lambda]_{\mathbf{L'}} \,\big|\, [x']_{\mathbf{M \times o'}}\big]^\dagger\Big] \quad (16)$$

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

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

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

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. (15) immediately reveals the marginal likelihood

$$\mathrm{p}\big([Y|E, F, \Lambda] \,\big|\, X\big) = \mathrm{p}\Big([Y]^\dagger_{\mathbf{L \times N}} \,\Big|\, [0]_{\mathbf{LN}}, K_Y\Big) \quad (17)$$

and the posterior distribution

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

The ultimate goal is the posterior predictive GP which extends the posterior distribution to arbitrary – usually unobserved – $[x]_{\mathbf{M \times o}}$. 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:

$$\big[[f_M|Y|E, F, \Lambda] \,\big|\, x\big]^\dagger_{\mathbf{L \times o}} \sim$$
$$\mathsf{N}\big(k(x, X)K_Y^{-1}Y^\dagger, \; k(x, x) - k(x, X)K_Y^{-1}k(X, x)\big) \quad (18)$$

4.4. GP Optimization

Henceforth we implicitly condition on optimal hyperparameters, which maximise the marginal likelihood Eq. (17).

$$[E]_{\mathbf{L^2}}, [F]_{\mathbf{L^2}}, [\Lambda]_{\mathbf{L \times M}} := \operatorname{argmax} p\left([Y]^{\dagger}_{\mathbf{L \times N}} \middle| [0]_{\mathbf{LN}}, K_Y \right) \qquad (19)$$

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

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

From this starting point, E, F may be optimized (off-diagonal elements in particular) in the full multi-output GP Eq. (19). One may then attempt to re-optimize lengthscales according to Eq. (19), and iterate, although this may be gilding the lily.

5. Gaussian Process Moments

This Section calculates the stochastic process moments of Gaussian Process Regression, absorbing Section 4 into the perspective of Section 3 Let $c \colon \mathbb{R} \to [0, 1]$ be the (bijective) CDF of the standard, univariate normal distribution, and define the triads

$$[z]_{\mathbf{M \times 3}} := c^{-1}\left([u]_{\mathbf{M \times 3}}\right) \sim \mathsf{N}\left([0]_{\mathbf{M \times 3}}, \langle 1 \rangle_{\mathbf{M^2}}\right)$$
$$[x]_{\mathbf{M' \times 3}} := [\Theta]^{\mathsf{T}}_{\mathbf{M \times M'}} [z]_{\mathbf{M \times 3}}$$

Here, the rotation matrix $[\Theta]^{\mathsf{T}}_{\mathbf{M \times M'}} = [\Theta]^{-1}_{\mathbf{M \times M'}}$ is broadcast to multiply the triad $[z]_{\mathbf{M \times 3}}$. 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.

Throughout the remainder of this paper, primed ordinal subscripts are used to specify Einstein sum (einsum) contraction of tensors. Whenever a subscript primed in a specific fashion appears in adjacent tensors (those not

separated by algebraic operations $+, -, \circ, \otimes$), it is einsummed over and does not subscript the result.

Adding shared gaussian noise $[\mathsf{e}_M | E]_\mathbf{L}$ to Eq. (18) yields

$$\left[y([\mathsf{u}]_{\mathbf{M+1\times 3}}) \middle| [u]_{\mathbf{M\times 3}} \right]^\dagger_{\mathbf{L\times 3}} = \left[[\mathsf{y}_\mathsf{M} | Y | E, F, \Lambda] \middle| [z]_{\mathbf{M\times 3}} \right]^\dagger_{\mathbf{L\times 3}} \sim$$
$$\mathsf{N}\left(k(x, X) K_Y^{-1} Y^\dagger, \ k(x, x) - k(x, X) K_Y^{-1} k(X, x) + E^\dagger \right) \quad (20)$$

using broadcast $[E^\dagger]_{\mathbf{L3\times L3}} := [[E]_{(\mathbf{L}\times 1)^2} \circ [1]_{(1\times \mathbf{3})^2}]^\dagger_{(\mathbf{L}\times \mathbf{3})^2}$. To bring the GP estimate fully under the umbrella of the SP estimate we should identify its ungovernable noise, and ascribe it to $[\mathsf{u}]_M$ of the SP. Let $d \colon (0,1) \to (0,1)^L$ concatenate every L^{th} decimal place starting at l, for each output dimension $l \leq L$ of $(0,1)^L$, then Eq. (20) can be written as

$$\left[y([\mathsf{u}]_{\mathbf{M+1\times 3}}) \middle| [u]_{\mathbf{M\times 3}} \right]^\dagger_{\mathbf{L\times 3}}$$
$$= [\mu_1]^\dagger_{\mathbf{L\times 3}} + [\mu_2]^{\dagger/2}_{\mathbf{L\times 3\times L'\times 3'}} \left[\left[c^{-1}(d\left([\mathsf{u}]_M\right)) \right]_{\mathbf{L}\times 1} \circ [1]_{1\times \mathbf{3}} \right]^\dagger_{\mathbf{L'\times 3'}} \quad (21)$$

where $[\mu_2]^{\dagger/2}_{(\mathbf{L\times 3})^2}$ denotes the Cholesky decomposition of the matrix $[\mu_2]^\dagger_{(\mathbf{L\times 3})^2}$. From the development in Section 3, the first two moments μ_1, μ_2 are sufficient to compute the coefficient of determination and its variance.

The crucial moments μ_1, μ_2 can be read straight off Eqs. (20) and (21), but still need conditioning. This is entirely a matter of repeatedly applying product formula Eq. (15), together with the familiar Gaussian identities

$$[\mathsf{z}]_\mathbf{M} \sim \mathsf{N}\left([Z]_\mathbf{M}, [\Sigma]_{\mathbf{M\times M}}\right) \Rightarrow [\mathsf{z}]_\mathbf{m} \sim \mathsf{N}\left([Z]_\mathbf{m}, [\Sigma]_{\mathbf{m\times m}}\right)$$
$$[\mathsf{z}]_\mathbf{m} \sim \mathsf{N}\left([Z]_\mathbf{m}, [\Sigma]_{\mathbf{m\times m}}\right) \Rightarrow [\Theta]^\mathsf{T}_{\mathbf{m\times m}} [\mathsf{z}]_\mathbf{m} \sim |\Theta|^{-1} \mathsf{N}(\Theta^\mathsf{T} Z, \Theta^\mathsf{T} \Sigma \Theta)$$

5.1. First Moments

The first moment of the GP for any $\mathbf{m} \subseteq \mathbf{M}$ is given by

$$[\mu_\mathbf{m}]_\mathbf{L} = \mathbb{E}_{\mathbf{M-m}}\left[k([\mathsf{x}]_\mathbf{M}, X) K_Y^{-1} Y^\dagger \middle| [z]_\mathbf{m} \right] = [g_\mathbf{m}]^\dagger_{\mathbf{L\times L'\times N'}} \left[K_Y^{-1} Y^\dagger \right]_{\mathbf{L'N'}}$$

where

$$[g_\mathbf{m}]_{l\times l'\times \mathbf{N'}} := [\pm F]_{l\times l'} \circ \mathrm{p}\left([0]_\mathbf{M} \middle| [X]_{\mathbf{M\times N'}}, \left\langle \Lambda^2 + 1 \right\rangle_{l\times l'} \right)$$
$$\circ \frac{\mathrm{p}\left([z]_\mathbf{m} \middle| [G]_{\mathbf{m}\times l\times l'\times \mathbf{N'}}, [\Gamma]_{l\times l'} \right)}{\mathrm{p}\left([z]_\mathbf{m} \middle| [0]_\mathbf{m}, \langle 1 \rangle_{\mathbf{m}^2} \right)}$$
$$= [\pm F]_{l\times l'} \circ \frac{\mathrm{p}\left([0]_\mathbf{M} \middle| [X]_{\mathbf{M\times N'}}, \left\langle \Lambda^2 + 1 \right\rangle_{l\times l'} \right)}{\mathrm{p}\left([0]_\mathbf{m} \middle| [G]_{\mathbf{m}\times l\times l'\times \mathbf{N'}}, [\Phi]_{l\times l'} \right)}$$
$$\circ \mathrm{p}\left([\Phi]_{l\times l'\times \mathbf{m\times m'}} [z]_{\mathbf{m'}} \middle| [G]_{\mathbf{m}\times l\times l'\times \mathbf{N'}}, [\Gamma]_{l\times l'} [\Phi]_{l\times l'\times \mathbf{m'\times m}} \right)$$

and

$$[G]_{\mathbf{m}\times l\times l'\times \mathbf{N}'} := [\Theta]_{\mathbf{m}\times \mathbf{M}} \left\langle \Lambda^2 + 1\right\rangle^{-1}_{l\times l'\times \mathbf{M}\times \mathbf{M}'} [X]_{\mathbf{M}'\times \mathbf{N}'}$$

$$[\Phi]_{l\times l'\times \mathbf{m}\times \mathbf{m}'} := [\Theta]_{\mathbf{m}\times \mathbf{M}} \left\langle \Lambda^2 + 1\right\rangle^{-1}_{l\times l'\times \mathbf{M}\times \mathbf{M}'} [\Theta]^{\mathsf{T}}_{\mathbf{m}'\times \mathbf{M}'}$$

$$[\Gamma]_{l\times l'\times \mathbf{m}^2} := \left\langle 1\right\rangle_{\mathbf{m}^2} - [\Phi]_{l\times l'\times \mathbf{m}^2}$$

In particular, when $\mathbf{m} = \mathbf{M}$, Θ factors out entirely. On the other hand

$$[g_0]_{\mathbf{L}^2\times \mathbf{N}'} = [\pm F]_{\mathbf{L}^2} \circ \mathrm{p}\left([0]_{\mathbf{M}} \,\middle|\, [X]_{\mathbf{M}\times \mathbf{N}'} , \left\langle \Lambda^2 + 1\right\rangle_{\mathbf{L}^2\times \mathbf{M}^2}\right)$$

Standardization of X and Y instills a totally marginal expectation of $\mu_0 \approx [0]_{\mathbf{L}}$, but this is usually inexact.

5.2. Second Moments

The second moment of the GP for any $\mathbf{m}, \mathbf{m}' \subseteq \mathbf{M}$ is given by

$$[\mu_{\mathbf{mm}'}]_{\mathbf{L}^2} = [F]_{\mathbf{L}^2} \circ [\phi_{\mathbf{mm}'}]_{\mathbf{L}^2} - [\psi_{\mathbf{mm}'}]_{\mathbf{L}^2} + [E]_{\mathbf{L}^2} \tag{22}$$

where

$$[\phi_{\mathbf{mm}'}]_{l\times l'} := \frac{\mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M}'-\mathbf{m}'}\left[k([\mathsf{x}]_{\mathbf{M}}, [\mathsf{x}]_{\mathbf{M}'}) \,\middle|\, [z]_{\mathbf{m}}, [z]_{\mathbf{m}'}\right]_{l\times l'}}{[F]_{l\times l'}}$$

$$= \frac{\left|\left\langle \Lambda^2\right\rangle_{l\times l'\times \mathbf{M}^2}\right|^{1/2} \mathrm{p}\left([\mathsf{z}]_{\mathbf{m}} \,\middle|\, [0]_{\mathbf{m}}, [\Upsilon]_{l\times l'\times \mathbf{m}^2}\right) \mathrm{p}\left([\mathsf{z}]_{\mathbf{m}'} \,\middle|\, [Z]_{l\times l'\times \mathbf{m}'}, [\Pi]_{l\times l'\times \mathbf{m}'^2}\right)}{\left|\left\langle \Lambda^2 + 2\right\rangle_{l\times l'\times \mathbf{M}^2}\right|^{1/2} \mathrm{p}\left([\mathsf{z}]_{\mathbf{m}} \,\middle|\, [0]_{\mathbf{m}}, \left\langle 1\right\rangle_{\mathbf{m}^2}\right) \mathrm{p}\left([\mathsf{z}]_{\mathbf{m}'} \,\middle|\, [0]_{\mathbf{m}'}, \left\langle 1\right\rangle_{\mathbf{m}'^2}\right)}$$

$$[\psi_{\mathbf{mm}'}]_{l\times l'} := \frac{\mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M}'-\mathbf{m}'}\left[k([\mathsf{x}]_{\mathbf{M}}, X)\, K_Y^{-1} k(X, [\mathsf{x}]_{\mathbf{M}'}) \,\middle|\, [z]_{\mathbf{m}}, [z]_{\mathbf{m}'}\right]_{l\times l'}}{[F]_{l\times l'}}$$

$$= [g_{\mathbf{m}}]^{\dagger}_{l\times \mathbf{L}''\times \mathbf{N}''} \left[K_Y^{-1}\right]_{\mathbf{L}''\mathbf{N}''\times \mathbf{L}'''\mathbf{N}'''} [g_{\mathbf{m}'}]^{\dagger}_{l'\times \mathbf{L}'''\times \mathbf{N}'''}$$

and

$$[\Upsilon]_{l\times l'\times \mathbf{m}\times \mathbf{m}''} := [\Theta]_{\mathbf{m}\times \mathbf{M}} \left\langle \left\langle \Lambda^2 + 1\right\rangle_{l\times l'} \left\langle \Lambda^2 + 2\right\rangle^{-1}_{l\times l'}\right\rangle_{\mathbf{M}\times \mathbf{M}'} [\Theta]^{\mathsf{T}}_{\mathbf{m}''\times \mathbf{M}'}$$

$$[\Pi]^{-1}_{l\times l'\times \mathbf{M}'\times \mathbf{M}'''} := \left\langle 1\right\rangle_{\mathbf{M}'\times \mathbf{M}'''} + [\Phi]_{l\times l'\times \mathbf{M}'\times \mathbf{M}'''} +$$
$$[\Phi]_{l\times l'\times \mathbf{M}'\times \mathbf{m}} [\Gamma]^{-1}_{l\times l'\times \mathbf{m}\times \mathbf{m}''} [\Phi]_{l\times l'\times \mathbf{m}''\times \mathbf{M}'''}$$

$$[Z]_{l\times l'\times \mathbf{m}'} := [\Pi]_{l\times l'\times \mathbf{m}'\times \mathbf{M}} [\Phi]_{l\times l'\times \mathbf{M}\times \mathbf{m}''} [\Gamma]^{-1}_{l\times l'\times \mathbf{m}''\times \mathbf{m}} [z]_{\mathbf{m}}$$

10

6. Gaussian Process Estimates

6.1. Expected Value

Using the shorthand

$$[KYg_0]_{l \times \mathbf{L'N'}} := [K_Y^{-1}Y^\dagger]_{\mathbf{L'N'}} \circ [g_0]^\dagger_{l \times \mathbf{L'} \times \mathbf{N'}}$$

to write

$$\mathbb{E}_\mathbf{m}\big[[\mu_\mathbf{m}]_\mathbf{L}^2\big]_{l \times l'} = [KYg_0]_{l \times \mathbf{L''N''}} [H_\mathbf{m}]^\dagger_{l \times \mathbf{L''} \times \mathbf{N''} \times l' \times \mathbf{L'''} \times \mathbf{N'''}} [KYg_0]_{l' \times \mathbf{L'''N'''}}$$

results in

$$[H_\mathbf{m}]_{l \times \mathbf{L''} \times \mathbf{N''} \times l' \times \mathbf{L'''} \times \mathbf{N'''}}$$

$$:= \mathbb{E}_\mathbf{m}\left[\frac{\mathrm{p}\big([z]_\mathbf{m}\big|\, [G]_{\mathbf{m} \times l \times \mathbf{L''} \times \mathbf{N''}}, [\Gamma]_{l \times \mathbf{L''}}\big) \otimes \mathrm{p}\big([z]_\mathbf{m}\big|\, [G]_{\mathbf{m} \times l' \times \mathbf{L'''} \times \mathbf{N'''}}, [\Gamma]_{l' \times \mathbf{L'''}}\big)}{\mathrm{p}\big([z]_\mathbf{m}\big|\, [0]_\mathbf{m}, \langle 1 \rangle_{\mathbf{m} \times \mathbf{m}}\big)\, \mathrm{p}\big([z]_\mathbf{m}\big|\, [0]_\mathbf{m}, \langle 1 \rangle_{\mathbf{m} \times \mathbf{m}}\big)} \right]$$

$$= \big[|\Psi|^{-1}\big]_{l \times \mathbf{L''} \times l' \times \mathbf{L'''}} \circ \frac{\mathrm{p}\big([G]_{\mathbf{m} \times l \times \mathbf{L''} \times \mathbf{N''}}\big|\, [G]_{\mathbf{m} \times l' \times \mathbf{L'''} \times \mathbf{N'''}}, [\Sigma]_{l \times \mathbf{L''} \times l' \times \mathbf{L'''}}\big)}{\mathrm{p}\big([0]_\mathbf{m}\big|\, [\Sigma G]_{\mathbf{m} \times l \times \mathbf{L''} \times \mathbf{N''} \times l' \times \mathbf{L'''} \times \mathbf{N'''}}, [\Sigma\Psi]_{l \times \mathbf{L''} \times l' \times \mathbf{L'''}}\big)}$$

where

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

$$[\Psi]_{l \times l'' \times l' \times l''' \times \mathbf{m} \times \mathbf{m}''} := [\Sigma]_{l \times l'' \times l' \times l''' \times \mathbf{m} \times \mathbf{m}''}$$
$$- [\Gamma]_{l \times l'' \times \mathbf{m} \times \mathbf{m}^{*'}} [\Gamma]_{l' \times l''' \times \mathbf{m}^{*'} \times \mathbf{m}''}$$

$$\big[|\Psi|^{-1}\big]_{l \times l'' \times l' \times l'''} := \big|[\Psi]_{l \times l'' \times l' \times l''' \times \mathbf{m}^2}\big|^{-1}$$

$$[\Sigma G]_{\mathbf{m} \times l \times l'' \times \mathbf{N''} \times l' \times l''' \times \mathbf{N'''}} := [\Gamma]_{l' \times l''' \times \mathbf{m} \times \mathbf{m}'} [G]_{\mathbf{m}' \times l \times l'' \times \mathbf{N''}}$$
$$+ [\Gamma]_{l \times l'' \times \mathbf{m} \times \mathbf{m}'} [G]_{\mathbf{m}' \times l' \times l''' \times \mathbf{N'''}}$$

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

6.2. Variance

In calculating

$$\mathbb{E}_\mathbf{m} \mathbb{E}_{\mathbf{m}'} [\mu_\mathbf{m} \otimes \mu_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2 \times \mathbf{L'}^2}$$

from Eq. (22) the terms containing E reduce to squares of g_0 by iterated expectations, and these will obviously cancel across Eq. (12). We may therefore

11

assume $E = 0$ in Eq. (22). The remaining terms are given by

$$\mathbb{E}_{\mathbf{m}} \mathbb{E}_{\mathbf{m}'} \left[[\mu_{\mathbf{m}}]_l \otimes [\phi_{\mathbf{mm}'}]_{l'' \times l'} \otimes [\mu_{\mathbf{m}'}]_{l'''} \right] =$$

$$\frac{\left| \langle \Lambda^2 \rangle_{l'' \times l' \times \mathbf{M}^2} \right|^{1/2} (2\pi)^{m/2}}{\left| \langle \Lambda^2 + 2 \rangle_{l'' \times l' \times \mathbf{M}^2} \right|^{1/2}} \left[K_Y^{-1} Y^\dagger \right]_{\mathbf{L}^* \mathbf{N}^*} \otimes \left[K_Y^{-1} Y^\dagger \right]_{\mathbf{L}^{*'} \mathbf{N}^{*'}}$$

$$\left[\frac{[g_0]_{l \times \mathbf{L}^* \times \mathbf{N}^*} \otimes [g_0]_{l''' \times \mathbf{L}^{*'} \times \mathbf{N}^{*'}}}{\mathrm{p}\left([0]_{\mathbf{m}'} \big| [G]_{\mathbf{m}' \times l''' \times \mathbf{L}^{*'} \times \mathbf{N}^{*'}}, [\Phi]_{l''' \times \mathbf{L}^{*'}} \right)} \circ \right.$$

$$\mathrm{p}\left([0]_{\mathbf{m}} \big| [1 - \Upsilon]_{l'' \times l'}^{1/2} [G]_{\mathbf{m} \times l \times \mathbf{L}^* \times \mathbf{N}^*}, \langle 1 \rangle + [1 - \Upsilon]_{l'' \times l'}^{1/2} [\Phi]_{l \times \mathbf{L}^*} [1 - \Upsilon]_{l'' \times l'}^{\mathsf{T}/2} \right) \circ$$

$$\left. \mathrm{p}\left([G]_{\mathbf{m}' \times l''' \times \mathbf{L}^{*'} \times \mathbf{N}^{*'}} \big| [\Omega] [C] [\Gamma]_{l \times \mathbf{L}^*}^{-1} [G]_{\mathbf{m} \times l \times \mathbf{L}^* \times \mathbf{N}^*}, [B] + [\Omega] [C] [\Omega]^{\mathsf{T}} \right) \right]^\dagger$$

using the Cholesky decomposition

$$\langle 1 \rangle_{\mathbf{m}^2} - [\Upsilon]_{l'' \times l' \times \mathbf{m}^2} = [1 - \Upsilon]_{l'' \times l'}^{1/2} [1 - \Upsilon]_{l'' \times l'}^{\mathsf{T}/2}$$

and

$$[\Omega]_{\mathbf{m}' \times \mathbf{m}} := [\Phi]_{l''' \times l^{*'} \times \mathbf{m}' \times \mathbf{m}'''} [\Pi]_{l'' \times l' \times \mathbf{m}''' \times \mathbf{M}} [\Phi]_{l'' \times l' \times \mathbf{M} \times \mathbf{m}''} [\Gamma]_{l'' \times l' \times \mathbf{m}'' \times \mathbf{m}}^{-1}$$

$$[B]_{\mathbf{m}'^2} := [\Gamma]_{l''' \times l^{*'} \times \mathbf{m}' \times \mathbf{m}^{*'}} [\Phi]_{l''' \times l^{*'} \times \mathbf{m}^{*'} \times \mathbf{m}'''} +$$

$$[\Phi]_{l''' \times l^{*'} \times \mathbf{m}' \times \mathbf{m}^{*'''}} [\Pi]_{l'' \times l' \times \mathbf{m}^{*'''} \times \mathbf{m}^{*'}} [\Phi]_{l''' \times l^{*'} \times \mathbf{m}^{*'} \times \mathbf{m}'''}$$

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

$$\left[\left[[\Gamma]_{l \times l^* \times \mathbf{m}^{*2}} + [\Phi]_{l \times l^* \times \mathbf{m}^{**} \times \mathbf{m}^{**'}} [\Upsilon]_{l'' \times l' \times \mathbf{m}^{**''} \times \mathbf{m}^{***}} \right]^{-1} \right]_{\mathbf{m}^* \times \mathbf{m}^{*''}}$$

$$[\Gamma]_{l \times l^* \times \mathbf{m}^{*''} \times \mathbf{m}''}$$

and

$$\mathbb{E}_{\mathbf{m}} \mathbb{E}_{\mathbf{m}'} \left[[\mu_{\mathbf{m}}]_l \otimes [\psi_{\mathbf{mm}'}]_{l'' \times l'} \otimes [\mu_{\mathbf{m}'}]_{l'''} \right] =$$

$$\left[K_Y^{-1} Y^\dagger \right]_{\mathbf{L}^* \mathbf{N}^*} \otimes \left[K_Y^{-1} \right]_{\mathbf{L}^{*''} \mathbf{N}^{*''} \times \mathbf{L}^{*'} \mathbf{N}^{*'}} \otimes \left[K_Y^{-1} Y^\dagger \right]_{\mathbf{L}^{*'''} \mathbf{N}^{*'''}}$$

$$\left[\frac{[g_0]_{l \times \mathbf{L}^* \times \mathbf{N}^*} \otimes [g_0]_{l'' \times \mathbf{L}^{*''} \times \mathbf{N}^{*''}} \otimes [g_0]_{l' \times \mathbf{L}^{*'} \times \mathbf{N}^{*'}} \otimes [g_0]_{l''' \times \mathbf{L}^{*'''} \times \mathbf{N}^{*'''}}}{\mathrm{p}\left([0]_{\mathbf{m}} \big| [G]_{\mathbf{m} \times l \times \mathbf{L}^* \times \mathbf{N}^*}, [\Phi]_{l \times \mathbf{L}^*} \right) \otimes \mathrm{p}\left([0]_{\mathbf{m}'} \big| [G]_{\mathbf{m}' \times l''' \times \mathbf{L}^{*'''} \times \mathbf{N}^{*'''}}, [\Phi]_{l''' \times \mathbf{L}^{*'''}} \right)} \right.$$

$$\circ \mathrm{p}\left([G]_{\mathbf{m} \times l \times \mathbf{L}^* \times \mathbf{N}^*} \big| [G]_{\mathbf{m} \times l'' \times \mathbf{L}^{*''} \times \mathbf{N}^{*''}}, [D]_{l \times \mathbf{L}^* \times l'' \times \mathbf{L}^{*''}} \right) \circ$$

$$\left. \mathrm{p}\left([G]_{\mathbf{m}' \times l' \times \mathbf{L}^{*'} \times \mathbf{N}^{*'}} \big| [G]_{\mathbf{m}' \times l''' \times \mathbf{L}^{*'''} \times \mathbf{N}^{*'''}}, [D]_{l' \times \mathbf{L}^{*'} \times l''' \times \mathbf{L}^{*'''}} \right) \right]^\dagger$$

where

$$[D]_{l' \times l^{*'} \times l''' \times l^{*'''} \times \mathbf{m}'^2} := [\Phi]_{l''' \times l^{*'''} \times \mathbf{m}' \times \mathbf{m}^{*'''}} -$$
$$[\Phi]_{l''' \times l^{*'''} \times \mathbf{m}' \times \mathbf{m}^{*'''}} [\Phi]_{l' \times l^{*'} \times \mathbf{m}^{*'''} \times \mathbf{m}^{*'}} [\Phi]_{l''' \times l^{*'''} \times \mathbf{m}^{*'} \times \mathbf{m}'''}$$

and similarly for $[D]_{l \times \mathbf{L}^* \times l'' \times \mathbf{L}^{*''} \times \mathbf{m}^2}$.

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`