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, Multi-Output, Gaussian Process

1. Introduction

This paper is concerned with analysing the results of experiments or computer simulations in a design matrix of $M \geq 1$ input and $L \geq 1$ output columns, over N rows (datapoints). Global Sensitivity Analysis [1] examines the relevance of the various inputs to the various outputs. When pursued via ANOVA decomposition, this leads naturally to the well known Sobol' indices, which have by now been applied across most fields of science and engineering [2, 3].

The Sobol' decomposition apportions the variance of the outputs to sets of one or more inputs [4]. We shall use ordinal sets of inputs $\mathbf{m} := (0, \ldots, m-1) \subseteq \mathbf{M}$, as tuples which are totally ordered sets. The maximal set $\mathbf{M}$ of all M inputs explains everything explicable, so its Sobol' index is 1 by definition. The void set $\mathbf{0}$ explains nothing, so its Sobol' index is 0 by definition. The influence of an isolated set of inputs $\mathbf{m}$ is measured by its closed Sobol' index $S_{\mathbf{m}} \in [0, 1]$. A first-order Sobol' index $S_{m'}$ is simply the closed Sobol' index of a single input m'. Because inputs in an isolated set may act in concert with

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)

each other, the influence of an isolated set often exceeds the sum of first-order contributions from its members, always obeying $S_{\mathbf{m}} \geq \sum_{m' \in \mathbf{m}} S_{m'}$.

The total Sobol index $S^T_{\mathbf{M}-\mathbf{m}} \geq 0$ of the set theoretic complement $\mathbf{M}-\mathbf{m}$ is $1-S_{\mathbf{m}}$, which expresses the influence of non-isolated inputs $\mathbf{M}-\mathbf{m}$ allowed to act in concert with each other *and* isolated inputs $\mathbf{m}$. When speaking of irrelevant inputs $\mathbf{M}-\mathbf{m}$, we mean that $S^T_{\mathbf{M}-\mathbf{m}} \approx 0$. This is synonymous with the isolated set of inputs $\mathbf{m}$ explaining everything explicable $S_{\mathbf{m}} \approx 1$. It is apparent that we can readily obtain any Sobol' index of interest by ordering input dimensions appropriately and calculating the closed index $S_{\mathbf{m}}$ of some ordinal set $\mathbf{m}$.

Apportioning variance is mathematically equivalent to squaring a correlation coefficient to produce a coefficient of determination R^2 [5]. A closed Sobol' index is thus a coefficient of determination between the predictions from a reduced model with $m \leq M$ inputs and predictions from the full model with M inputs. Simplicity and economy (not least of calculation) motivate the adoption of a reduced model, a closed Sobol' index close to 1 permits it. Why on earth would one use the full model, when its predictions are almost identical to the reduced model?

With multiple outputs, the Sobol' decomposition apportions the covariance matrix of outputs [6], rather than the variance of a single output. With L outputs, the closed Sobol' index $S_{\mathbf{m}}$ is a symmetric $L \times L$ matrix. The diagonal elements express the relevance of inputs to the output variables themselves. The off-diagonal elements express relevance to the linkages between outputs. This may be of considerable interest when outputs are, for example, yield and purity of a product, or perhaps a single output measured at various times. The Sobol indices reveal (amongst other things) which inputs it is worthwhile varying in an effort to alter the linkages between outputs.

Accurate calculation of Sobol' indices even for a single output is computationally expensive and requires 10,000+ datapoints [7]. A more efficient approach is calculation via a surrogate model, such as Polynomial Chaos Expansion [8, 9, 10], low-rank tensor approximation [11, 12], and support vector regression [13]. As well as being efficient, surrogate models also smooth out noise in the outputs, which is often highly desirable in practice. This paper employs one of the most popular surrogates, the Gaussian Processes (GP) [14, 15] as it is highly tractable. We shall follow the multi-output form (MOGP) described in [16], in order to examine the linkages between outputs.

Semi-analytic expressions for Sobol' indices have been provided in integral form by [17] and alternatively by [18]. These approaches are implemented, examined and compared in [19, 20]. Both [17, 19] estimate the errors on Sobol' indices in semi-analytic, integral form. Fully analytic, closed form expressions have been derived without error estimates for uniformly distributed inputs [21] with an RBF kernel. There are currently no closed form expressions for MOGPs, or the errors on Sobol' indices, or any GPs for which inputs are not uniformly distributed.

In this paper we provide explicit, closed-form analytic formulae for closed Sobol' indices and their errors, for a class of MOGP with an anisotropic radial basis function (RBF/ARD) kernel applicable to smoothly varying outputs. We transform uniformly distributed inputs u to normally distributed inputs z prior to fitting a GP and performing analytic calculation of closd Sobol' indices. This leads to relatively concise expressions in terms of exponentials, and enables ready calculation of the errors (variances) of these expressions. It also allows for an arbitrary rotation Θ of inputs, as normal variables are additive, whereas summing uniform inputs does not produce uniform inputs. If the goal is reducing inputs, rotating their basis first boosts the possibilities immensely [22]. It presents the possibilty of choosing Θ to maximise the closed Sobol' index of the first few inputs.

The quantities to be calculated and their formal context are introduced in Section 2. Our approach effectively regards a regression model furnishing an uncertainty measure with each prediction as just another name for a stochastic process. A great deal of progress is made in Section 3 using general stochastic (not necessarily Gaussian) processes. This approach is analytically cleaner, as it is not obfuscated by the GP details. Furthermore, it turns out that the desirable properties of the Gaussian (lack of skew, simple kurtosis) are not actually helpful, as these terms cancel of their own accord. This development leaves just two terms to be calculated, which require the stochastic process to be specified. MOGPs with an RBF/ARD kernel are tersely developed and described in Section 4, then used to calculate the two unknown terms in Sections 5 and 6. Conclusions are drawn in Section 7.

2. Coefficient of Determination

Given a model

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

take as input a uniformly distributed random variable (RV)

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

Throughout this paper exponentiation is categorical – repeated cartesian $\times$ or tensor $\otimes$ – unless otherwise specified. Square bracketed quantities are tensors, carrying their axes as a subscript tuple. In this case the subscript tuple is the von Neumann ordinal

$$\mathbf{M}+\mathbf{1} := (0,\ldots,M) \supset \mathbf{m} := (0,\ldots,m-1 \leq M-1)$$

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

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

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

ranging from $[\mathsf{y}_0]_{\mathbf{L}}$ to $[\mathsf{y}_\mathsf{M}]_{\mathbf{L}}$. Every SP depends stochastically on the ungovernable noise dimension $[\mathsf{u}]_M \perp [\mathsf{u}]_{\mathbf{M}}$ and deterministically on the first m governed inputs $[u]_{\mathbf{m}}$, marginalizing the remaining inputs $[\mathsf{u}]_{\mathbf{M}-\mathbf{m}}$. Sans serif symbols such as u, y generally refer to RVs and SPs, italic u, y being reserved for (tensor) functions and variables. Each SP is simply a regression model for y on the first m dimensions of u.

Following the Kolmogorov extension theorem [23] 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 $o \in \mathbb{Z}^+$ input samples. The Kolmogorov extension theorem 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 $n \in \mathbb{Z}^+$ [24]. So Fubini's Theorem [25] pp.77 allows all expectations to be taken in any order. These observations suffice to ensure every object appearing in this work. Changing the order of expectations, as permitted by Fubini's Theorem, is the vital tool in the construction of this work.

Our aim is to compare predictions from a reduced regression model $\mathsf{y_m}$ with those from the full regression 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 or closed Sobol' index

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

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

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

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

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

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

In practice it is best to retain only the term in $W_{\mathbf{mm}}$, ignoring the uncertainty in $V_{\mathbf{M}}$ conveyed by $W_{\mathbf{Mm}}, W_{\mathbf{MM}}$. This is because one is normally interested in adequate reduced models, for which $V_{\mathbf{m}} \approx V_{\mathbf{M}}$ implies $W_{\mathbf{mm}} - 2W_{\mathbf{Mm}} + W_{\mathbf{MM}} \approx 0$, yielding a drastically vanishing uncertainty in the Sobol' index.

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

3. Stochastic Process Estimates

The central problem in calculating errors on Sobol' indices is that they involve ineluctable covariances between differently marginalized SPs, via their moments over ungovernable noise. But marginalization and moment determination are both a matter of taking expectations. So the ineluctable can be avoided by reversing the order of expectations – taking moments over ungovernable noise, then marginalizing. To this end, adopt as design matrix a triad of inputs to condition $[\mathsf{u}]_{\mathbf{M+1\times 3}}$, eliciting the response

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

Primes mark independent inputs, otherwise expectations are shared by all three members of the triad. It is not always obvious whether inputs 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 by the triad members)

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

for these embody

$$[\mu_{\mathbf{m'}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := [\mu_n]_{\prod_{j=1}^n(\mathbf{L}\times i_j)} = \mathbb{E}_M[[\mathsf{y}_{\mathbf{m'}}]_{\mathbf{L}}\otimes\cdots\otimes[\mathsf{y}_{\mathbf{m}^{n\prime}}]_{\mathbf{L}}]$$

where $i_j\in\mathbf{3}$ corresponds to $\mathbf{m}^{j\prime}\in\{\mathbf{0},\mathbf{m},\mathbf{M}\}$. This expression underpins the quantities we seek. The reduction which follows repeatedly realizes

$$[\mu_{\mathbf{0}\ldots\mathbf{0}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := \mathbb{E}_{\mathbf{M}}[\mu_{\mathbf{M}\ldots\mathbf{M}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} = \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}\ldots\mathbf{m}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} \tag{9}$$

Defining

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

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

$$\begin{aligned}[V_{\mathbf{m}}]_{\mathbf{L}^2} &= \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_M\left[[\mathsf{e}_{\mathsf{m}}+\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right]-\mathbb{E}_M\left[[\mathsf{e}_0+\mu_{\mathbf{0}}]^2_{\mathbf{L}}\right]\\ &= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right]-[\mu_{\mathbf{0}}]^2_{\mathbf{L}}+\mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{mm}}]_{\mathbf{L}^2}-[\mu_{\mathbf{00}}]_{\mathbf{L}^2}\\ &= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right]-[\mu_{\mathbf{0}}]^2_{\mathbf{L}}\end{aligned} \tag{11}$$

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

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

Here, the inputs within any $\mathbf{m}, \mathbf{m}' \subseteq \mathbf{M}$ clearly vary independently, and

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

exploiting the fact that $V_{\mathbf{0}} = [0]_{\mathbf{L}^2}$. Equation (9) cancels all terms beginning with $[\mathsf{e}_{\mathsf{m}}]^2_{\mathbf{L}}$, first across $A_{\mathbf{mm}'} - A_{\mathbf{0m}'}$ then across $A_{\mathbf{m0}} - A_{\mathbf{00}}$. All remaining terms ending in $[\mu_{\mathbf{m}'}]^2_{\mathbf{L}}$ are eliminated by centralization $\mathbb{E}_M[\mathsf{e}_{\mathsf{m}}] = 0$ and

$$\begin{aligned}&\mathbb{E}_{\mathbf{m}} \, \mathbb{E}_{\mathbf{m}'} \left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [\mu_{\mathbf{m}'}]^2_{\mathbf{L}} \right] - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}'}]_{\mathbf{L}^2} \\ &\qquad\qquad = [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [\mu_{\mathbf{0}}]^2_{\mathbf{L}} + [\mu_{\mathbf{0}}]^2_{\mathbf{L}} \otimes [V_{\mathbf{m}'}]_{\mathbf{L}^2} + [\mu_{\mathbf{0}}]^4_{\mathbf{L}}\end{aligned}$$

cancelling across $A_{\mathbf{mm}'} - A_{\mathbf{0m}'} - A_{\mathbf{m0}} + A_{\mathbf{00}}$. Similar arguments eliminate $[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}}$ and $[\mu_{\mathbf{m}}]^2_{\mathbf{L}}$. Effectively then

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

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 ($\mathbf{L}^{j\prime} = \mathbf{L}$ always), they do not change the value of the constant – unlike primes on variables ($\mathbf{m}^{j\prime}$ need not equal $\mathbf{m}$ in general). One is normally only interested in variances (errors), constituted by the diagonal $\mathbf{L}'^2 = \mathbf{L}^2$, for which the summation in Eq. (13) is over a pair of identical pairs.

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. Then all the answers we desire are given by Eqs. (3) and (11), and Eqs. (5), (12) and (13).

4. Interlude: Gaussian Process Regression

The development in this Section is based on [16], with slightly different notation. A Gaussian Process (GP) over x is formally defined and specified by

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

where tensor ranks concatenate into a multivariate normal distribution

$$[]_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}^{\dagger}\Big([]_{\mathbf{L}\times\mathbf{o}}\,,[]_{(\mathbf{L}\times\mathbf{o})^2}\Big) \Longleftrightarrow []^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\Big([]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\,,[]^{\dagger}_{(\mathbf{L}\times\mathbf{o})^2}\Big)$$
$$\Big[[]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\Big]_{\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o}} := []_{(l-1)\times\mathbf{o}}$$
$$\Big[[]^{\dagger}_{(\mathbf{L}\times\mathbf{o})^2}\Big]_{(\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o})\times(\mathbf{l'o}-(\mathbf{l'}-\mathbf{1})\mathbf{o})} := []_{(l-1)\times\mathbf{o}\times(l'-1)\times\mathbf{o}}$$

supporting the fundamental definition of the GP kernel, as a covariance (over ungovernable noise) between responses

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

4.1. Tensor Gaussians

Henceforth, tensors will be broadcast when necessary, as described in [26, 27]. This means that ranks and dimensions are implicitly expanded as necessary to perform an algebraic operation between tensors of differing signature. A tensor Gaussian like $\mathrm{p}\big([x]_{\mathbf{m}\times\mathbf{o}}\big|\,[x']_{\mathbf{m}\times\mathbf{o'}}\,,[\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\big)$ is defined element-wise, using broadcasting

$$\big[\mathrm{p}\big([x]_{\mathbf{m}\times\mathbf{o}}\big|\,[x']_{\mathbf{m}\times\mathbf{o'}}\,,[\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\big)\big]_{l\times o\times l'\times o'} := (2\pi)^{-M/2}\left|[\Sigma]_{l\times l'}\right|^{-1/2}$$
$$\exp\left(-\frac{[x-x']^{\intercal}_{\mathbf{m}\times l\times o\times l'\times o'}\,[\Sigma]^{-1}_{l\times l'\times\mathbf{m}\times\mathbf{m'}}\,[x-x']_{\mathbf{m'}\times l\times o\times l'\times o'}}{2}\right) \quad (14)$$

for $\mathbf{m'}=\mathbf{m}$ and transposition $^{\intercal}$ moving first rank to last.

Remarkably, the algebraic development in the remainder of this paper relies almost exclusively on an invaluable product formula reported in [28]:

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

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

4.2. Prior GP

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

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

Angle brackets denote a (perhaps broadcast) diagonal tensor, such as the identity matrix $\langle 1\rangle_{(1\times\mathbf{o})^2} =: \big\langle [1]_{(1\times\mathbf{o})^2}\big\rangle$.

The RBF kernel is hyperparametrized by signal covariance $[F]_{\mathbf{L}^2}$ and the tensor $[\Lambda]_{\mathbf{L}^2\times\mathbf{M}}$ of characteristic lengthscales, which must be symmetric $[\Lambda]_{l\times l'\times\mathbf{M}} = [\Lambda]_{l'\times l\times\mathbf{M}}$. Now use

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

to implement the non-informative RBF prior according to Eq. (14)

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

4.3. Predictive GP

Bayesian inference for GP regression further conditions the hyper-parametrized GP $\mathsf{y}|E,F,\Lambda$ on the observed realization (over ungovernable noise) 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

$$\begin{aligned}
[K_{\mathsf{e}}]_{\mathbf{Lo}\times\mathbf{Lo}} &:= \mathbb{V}_M\Big[\Big[[\mathsf{e}_M|E]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\Big]^{\dagger}\Big]\\
&= \Big[[E]_{(\mathbf{L}\times 1)^2}\circ\langle 1\rangle_{(1\times\mathbf{o})^2}\Big]^{\dagger}\\
[k(x,x')]_{\mathbf{Lo}\times\mathbf{Lo'}} &:= \mathbb{V}_M\Big[\Big[[\mathsf{f_M}|F,\Lambda]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\Big]^{\dagger}, \Big[[\mathsf{f_M}|F,\Lambda]_{\mathbf{L}}\,\big|\,[x']_{\mathbf{M}\times\mathbf{o'}}\Big]^{\dagger}\Big]\\
&= \Big[[\pm F]_{\mathbf{L}^2}\circ\mathrm{p}\Big([x]_{\mathbf{M}\times\mathbf{o}}\big|\,[x']_{\mathbf{M}\times\mathbf{o'}}\,,\big\langle\Lambda^2\big\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\Big)\Big]^{\dagger}\\
[K_Y]_{\mathbf{LN}\times\mathbf{LN}} &:= \mathbb{V}_M\Big[\Big[[\mathsf{y}|E,F,\Lambda]_{\mathbf{L}}\,\big|\,[X]_{\mathbf{M}\times\mathbf{N}}\Big]^{\dagger}\Big]\\
&= k([X]_{\mathbf{M}\times\mathbf{N}}\,,[X]_{\mathbf{M}\times\mathbf{N}}) + [K_{\mathsf{e}}]_{\mathbf{LN}\times\mathbf{LN}}
\end{aligned}\tag{16}$$

Applying Bayes' rule

$$\begin{aligned}
\mathsf{p}(\mathsf{f_M}|Y)\mathsf{p}(Y) = \mathsf{p}(Y|\mathsf{f_M})\mathsf{p}(\mathsf{f_M}) &= \mathrm{p}\big(Y^{\dagger}\big|\,\mathsf{f_M}^{\dagger},K_{\mathsf{e}}\big)\,\mathrm{p}\big(\mathsf{f_M}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,,k(X,X)\big)\\
&= \mathrm{p}\big(\mathsf{f_M}^{\dagger}\big|\,Y^{\dagger},K_{\mathsf{e}}\big)\,\mathrm{p}\big(\mathsf{f_M}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,,k(X,X)\big)
\end{aligned}$$

Product formula Eq. (15) immediately reveals the marginal likelihood

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

and the posterior distribution

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

The ultimate goal is the posterior predictive GP which extends the posterior distribution to arbitrary – usually unobserved – $[x]_{\mathbf{M}\times\mathbf{o}}$. This is formally derived from the definition of conditional probability, but this seems unnecessary, for the extension must recover the posterior distribution when $x = X$. Without unfeasible distortions, there is only one way of selectively replacing X with x in the posterior formula which preserves the coherence of tensor ranks:

$$\big[[\mathsf{f}_{\mathsf{M}}|Y|E,F,\Lambda]\,\big|x\big]^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x)-k(x,X)K_Y^{-1}k(X,x)\big) \tag{18}$$

In order to calculate the last term, the Cholesky decomposition $K_Y^{1/2}$ is used to write

$$[k(x,X)K_Y^{-1}k(X,x)]_{\mathbf{Lo}^2} = [K_Y^{-1/2}k(X,x)]^2_{\mathbf{Lo}}$$

4.4. GP Optimization

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

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

5. Gaussian Process Moments

This Section calculates the stochastic process moments of GP 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

$$[\mathsf{z}]_{\mathbf{M}\times\mathbf{3}} := c^{-1}\big([\mathsf{u}]_{\mathbf{M}\times\mathbf{3}}\big) \sim \mathsf{N}\big([0]_{\mathbf{M}\times\mathbf{3}}\,, \langle 1\rangle_{\mathbf{M}^2}\big)$$
$$[\mathsf{x}]_{\mathbf{M}'\times\mathbf{3}} := [\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}'}\,[\mathsf{z}]_{\mathbf{M}\times\mathbf{3}}$$

Here, the rotation matrix $[\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}'} = [\Theta]^{-1}_{\mathbf{M}\times\mathbf{M}'}$ is broadcast to multiply the triad $[\mathsf{z}]_{\mathbf{M}\times\mathbf{3}}$. The purpose of ths arbitrary rotation is to allow GPs whose input basis x is not aligned with the fundamental basis u of the coefficient of determination. The latter is aligned with z which is the input we must condition. This generalization is cheap, given product formula Eq. (15), and of great potential benefit. One could, for example, imagine optimizing Θ to maximize $S_{\mathbf{m}}$.

Throughout the remainder of this paper, primed ordinal subscripts are used to specify Einstein sum (einsum) contraction of tensors, the multiplication and summation of elements over a matching index which underpins matrix multiplication. In this work, whenever a subscript primed in a specific fashion appears in adjacent tensors (those not separated by algebraic operations $+, -, \circ, \otimes$) and does not subscript the result, it is einsummed over. Detailed examples of the convention are given under `einsum` in [26].

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

$$\left[y([\mathsf{u}]_{\mathbf{M+1}\times\mathbf{3}}) \middle| \, [u]_{\mathbf{M}\times\mathbf{3}}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} = \left[[\mathsf{y}_{\mathsf{M}}|Y|E,F,\Lambda] \, \middle| [z]_{\mathbf{M}\times\mathbf{3}}\right]^{\dagger}_{\mathbf{L}\times\mathbf{3}} \sim \mathsf{N}\Big(k(x,X)K_Y^{-1}Y^{\dagger}, \; k(x,x) - [K_Y^{-1/2}k(X,x)]^2_{\mathbf{Lo}} + E^{\dagger}\Big) \quad (20)$$

using broadcast $[E^{\dagger}]_{\mathbf{L3}\times\mathbf{L3}} := [[E]_{(\mathbf{L}\times 1)^2} \circ [1]_{(1\times\mathbf{3})^2}]^{\dagger}_{(\mathbf{L}\times\mathbf{3})^2}$. 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

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

where $[\mu_2]^{\dagger/2}_{(\mathbf{L}\times\mathbf{3})^2}$ denotes the lower triangular Cholesky decomposition of the matrix $[\mu_2]^{\dagger}_{(\mathbf{L}\times\mathbf{3})^2}$. 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 are simply read from 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

$$\begin{aligned}[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}\,,[\Sigma]_{\mathbf{M}\times\mathbf{M}}\big) &\Rightarrow [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}\,,[\Sigma]_{\mathbf{m}\times\mathbf{m}}\big)\\ [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}\,,[\Sigma]_{\mathbf{m}\times\mathbf{m}}\big) &\Rightarrow [\Theta]^{\intercal}_{\mathbf{m}\times\mathbf{m}}\,[\mathsf{z}]_{\mathbf{m}} \sim |\Theta|^{-1}\,\mathsf{N}(\Theta^{\intercal}Z,\Theta^{\intercal}\Sigma\Theta)\end{aligned}$$

Henceforth the ordinal set $\mathbf{m}''$, whether or not decorated with a further *even* number of primes, should be taken as equal to $\mathbf{m}$. Likewise the ordinal set $\mathbf{m}'''$, whether or not decorated with a further *even* number of primes, should be taken as equal to $\mathbf{m}'$. Superscript * will stand for four consecutive primes $''''$. So $\mathbf{m},\mathbf{m}'$ are identified by the parity (even or odd) of the primes adorning $\mathbf{m}$. Such fussy ornamentation is necessary to maintain the integrity of einstein summation. This encumbrance applies to ordinal sets, not singleton values, so the many different prime decorations of l *always* indicate potentially different values.

5.1. First Moments

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

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

where

$$\begin{aligned}\frac{[g_{\mathbf{m}}]_{l\times l''\times\mathbf{N}''}}{[g_{\mathbf{0}}]_{l\times l''\times\mathbf{N}''}} &:= \frac{\mathrm{p}\big([z]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Gamma]_{l\times l''}\big)}{\mathrm{p}([z]_{\mathbf{m}}|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}^2})}\\ &= \frac{\mathrm{p}\big([\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}}\,[z]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}''}\,[\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}}\big)}{\mathrm{p}\big([0]_{\mathbf{m}}|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Phi]_{l\times l''}\big)}\end{aligned}$$

and

$$\begin{aligned}[g_{\mathbf{0}}]_{l\times l''\times\mathbf{N}''} &:= [\pm F]_{l\times l''}\,\mathrm{p}\Big([0]_{\mathbf{M}}|\,[X]_{\mathbf{M}\times\mathbf{N}''}\,,\big\langle\Lambda^2+1\big\rangle_{l\times l''}\Big)\\ [G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}}\,\big\langle\Lambda^2+1\big\rangle^{-1}_{l\times l''\times\mathbf{M}\times\mathbf{M}''}\,[X]_{\mathbf{M}''\times\mathbf{N}''}\\ [\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}} &:= [\Theta]_{\mathbf{m}''\times\mathbf{M}}\,\big\langle\Lambda^2+1\big\rangle^{-1}_{l\times l''\times\mathbf{M}\times\mathbf{M}''}\,[\Theta]^{\intercal}_{\mathbf{m}\times\mathbf{M}''}\\ [\Gamma]_{l\times l''\times\mathbf{m}^2} &:= \langle 1\rangle_{\mathbf{m}^2} - [\Phi]_{l\times l''\times\mathbf{m}^2}\end{aligned}$$

Note that when $\mathbf{m}=\mathbf{M}$, Θ factors out entirely. The unconditional expectation $\mu_{\mathbf{0}} \approx \big[\bar{Y}\big]_{\mathbf{L}}$, 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

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

where

$$[\phi_{\mathbf{m}\mathbf{m}'}]_{l\times l'} := \frac{\mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M}'-\mathbf{m}'}\big[k([\mathsf{x}]_{\mathbf{M}}\,,[\mathsf{x}]_{\mathbf{M}'})\,\big|\,[z]_{\mathbf{m}}\,,[z]_{\mathbf{m}'}\big]_{l\times l'}}{[F]_{l\times l'}}$$
$$= \frac{\big|\langle\Lambda^2\rangle_{l\times l'\times\mathbf{M}^2}\big|^{1/2}\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[0]_{\mathbf{m}}\,,[1-\Upsilon]_{l\times l'\times\mathbf{m}^2}\big)\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}'}\big|\,[Z]_{l\times l'\times\mathbf{m}'}\,,[\Pi]_{l\times l'\times\mathbf{m}'^2}\big)}{\big|\langle\Lambda^2+2\rangle_{l\times l'\times\mathbf{M}^2}\big|^{1/2}\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}^2}\big)\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}'}\big|\,[0]_{\mathbf{m}'}\,,\langle 1\rangle_{\mathbf{m}'^2}\big)}$$

$$\begin{aligned}[\psi_{\mathbf{m}\mathbf{m}'}]_{\mathbf{L}\times\mathbf{L}'} &:= \mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M}'-\mathbf{m}'}\big[k([\mathsf{x}]_{\mathbf{M}}\,,X)\,K_Y^{-1}k(X,[\mathsf{x}]_{\mathbf{M}'})\,\big|\,[z]_{\mathbf{m}}\,,[z]_{\mathbf{m}'}\big]_{\mathbf{L}\times\mathbf{L}'}\\ &= \Big([g_{\mathbf{m}}]^{\dagger}_{\mathbf{L}\times\mathbf{L}''\times\mathbf{N}''}\,[K_Y]^{-1/2}_{\mathbf{L}'''\mathbf{N}'''\times\mathbf{L}''\mathbf{N}''}\Big)\,\Big([g_{\mathbf{m}'}]^{\dagger}_{\mathbf{L}'\times\mathbf{L}''\times\mathbf{N}''}\,[K_Y]^{-1/2}_{\mathbf{L}'''\mathbf{N}'''\times\mathbf{L}''\mathbf{N}''}\Big)\end{aligned}$$

using the lower triangular Cholesky decomposition $[K_Y]^{1/2}_{\mathbf{LN}\times\mathbf{LN}}$ and

$$\begin{aligned}[\Upsilon]_{l\times l'\times\mathbf{m}\times\mathbf{m}''} &:= [\Phi]_{l\times l'\times\mathbf{m}\times\mathbf{M}}\,\big\langle\Lambda^2+2\big\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}'}\,[\Phi]^{\intercal}_{l\times l'\times\mathbf{m}''\times\mathbf{M}'}\\ [\Pi]^{-1}_{l\times l'\times\mathbf{M}'\times\mathbf{M}'''} &:= \langle 1\rangle_{\mathbf{M}'\times\mathbf{M}'''} + [\Phi]_{l\times l'\times\mathbf{M}'\times\mathbf{M}'''} +\\ &\qquad [\Phi]_{l\times l'\times\mathbf{M}'\times\mathbf{m}}\,[\Gamma]^{-1}_{l\times l'\times\mathbf{m}\times\mathbf{m}''}\,[\Phi]_{l\times l'\times\mathbf{m}''\times\mathbf{M}'''}\\ [Z]_{l\times l'\times\mathbf{m}'} &:= [\Pi]_{l\times l'\times\mathbf{m}'\times\mathbf{M}}\,[\Phi]_{l\times l'\times\mathbf{M}\times\mathbf{m}''}\,[\Gamma]^{-1}_{l\times l'\times\mathbf{m}''\times\mathbf{m}}\,[\mathsf{z}]_{\mathbf{m}}\end{aligned}$$

6. Gaussian Process Estimates

6.1. Expected Value

Using the shorthand

$$[g_{\mathbf{0}}KY]^{\dagger}_{l\times\mathbf{L}''\mathbf{N}''} := [g_{\mathbf{0}}]^{\dagger}_{l\times\mathbf{L}''\times\mathbf{N}''} \circ \big[K_Y^{-1}Y^{\dagger}\big]_{\mathbf{L}''\mathbf{N}''}$$

to write

$$\mathbb{E}_{\mathbf{m}}\big[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\big]_{l\times l'} =: [g_{\mathbf{0}}KY]^{\dagger}_{l\times\mathbf{L}''\mathbf{N}''}\,[H_{\mathbf{m}}]^{\dagger}_{l\times\mathbf{L}''\times\mathbf{N}''\times l'\times\mathbf{L}'''\times\mathbf{N}'''}\,[g_{\mathbf{0}}KY]^{\dagger}_{l'\times\mathbf{L}'''\mathbf{N}'''}$$

results in

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

where

$$
\begin{aligned}
[\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}^*\times\mathbf{m}''}:=&\,[\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}''}+[\Gamma]_{l'\times l'''\times\mathbf{m}^*\times\mathbf{m}''}\\
&-[\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}}\,[\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}''}
\end{aligned}
$$

6.2. Variance

Recall from Eq. (12) that the inputs comprising $\mathbf{m},\mathbf{m}'$ vary independently when calculating a covariance $W_{\mathbf{m}\mathbf{m}'}$ via $A_{\mathbf{m}\mathbf{m}'}$. In calculating

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

in Eq. (22) the terms containing the ungovernable noise variance $[E]_{\mathbf{L}^2}$ reduce to the same function of $g_{\mathbf{0}}$ by reduction formula Eq. (9), so these will obviously cancel across the four $A_{\mathbf{m}\mathbf{m}'}$ terms in Eq. (12). We may therefore assume $E=0$ in Eq. (22). This leaves just two terms, which we report again using superscript * to stand for four consecutive primes $''''$. Firstly

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m}'}\big[[\mu_{\mathbf{m}}]_l\otimes[\phi_{\mathbf{m}\mathbf{m}'}]_{l''\times l'''}\otimes[\mu_{\mathbf{m}'}]_{l'}\big]=\\
&\qquad\frac{\big|\langle\Lambda^2\rangle_{l''\times l'''\times\mathbf{M}^2}\big|^{1/2}(2\pi)^{m/2}}{\big|\langle\Lambda^2+2\rangle_{l''\times l'''\times\mathbf{M}^2}\big|^{1/2}}\,[g_{\mathbf{0}}KY]_{l\times\mathbf{L}^*\mathbf{N}^*}\otimes[g_{\mathbf{0}}KY]_{l'\times\mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}\\
&\qquad\Bigg[\mathrm{p}\Big([0]_{\mathbf{m}}\big|\,[\Upsilon]^{1/2}_{l''\times l'''}\,[G]_{\mathbf{m}\times l\times\mathbf{L}^*\times\mathbf{N}^*},\langle 1\rangle-[\Upsilon]^{1/2}_{l''\times l'''}\,[\Phi]_{l\times\mathbf{L}^*}\,[\Upsilon]^{\intercal/2}_{l''\times l'''}\Big)\\
&\qquad\circ\frac{\mathrm{p}\big([G]_{\mathbf{m}'\times l'\times\mathbf{L}^{*\prime}\times\mathbf{N}^{*\prime}}\big|\,[\Omega]\,[C]\,[\Gamma]^{-1}_{l\times\mathbf{L}^*}\,[G]_{\mathbf{m}\times l\times\mathbf{L}^*\times\mathbf{N}^*},[B]+[\Omega]\,[C]\,[\Omega]^\intercal\big)}{\mathrm{p}\big([0]_{\mathbf{m}'}\big|\,[G]_{\mathbf{m}'\times l'\times\mathbf{L}^{*\prime}\times\mathbf{N}^{*\prime}},[\Phi]_{l'\times\mathbf{L}^{*\prime}}\big)}\Bigg]^\dagger
\end{aligned}
$$

using the lower triangular Cholesky decomposition

$$
[\Upsilon]_{l''\times l'''\times\mathbf{m}^2}=[\Upsilon]^{1/2}_{l''\times l'''}\,[\Upsilon]^{\intercal/2}_{l''\times l'''}
$$

and

$$
\begin{aligned}
[\Omega]_{\mathbf{m}'\times\mathbf{m}} &:= [\Phi]_{l'\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}'''}\,[\Pi]_{l''\times l'''\times\mathbf{m}'''\times\mathbf{M}}\,[\Phi]_{l''\times l'''\times\mathbf{M}\times\mathbf{m}''}\,[\Gamma]^{-1}_{l''\times l'''\times\mathbf{m}''\times\mathbf{m}} \\
[B]_{\mathbf{m}'\times\mathbf{m}'''} &:= [\Gamma]_{l'\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime}}\,[\Phi]_{l'\times l^{*\prime}\times\mathbf{m}^{*\prime}\times\mathbf{m}'''} + \\
&\quad\ [\Phi]_{l'\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime\prime\prime}}\,[\Pi]_{l''\times l'''\times\mathbf{m}^{*\prime\prime\prime}\times\mathbf{m}^{*\prime}}\,[\Phi]_{l'\times l^{*\prime}\times\mathbf{m}^{*\prime}\times\mathbf{m}'''} \\
[C]_{\mathbf{m}\times\mathbf{m}''} &:= [1-\Upsilon]_{l''\times l'''\times\mathbf{m}\times\mathbf{m}^{*}} \\
&\quad\ \big[\langle 1\rangle - [\Phi]_{l\times l^{*}\times\mathbf{m}^{**}\times\mathbf{m}^{**\prime\prime}}\,[\Upsilon]_{l''\times l'''\times\mathbf{m}^{**\prime\prime}\times\mathbf{m}^{***}}\big]^{-1}_{\mathbf{m}^{*}\times\mathbf{m}^{*\prime\prime}} \\
&\quad\ [\Gamma]_{l\times l^{*}\times\mathbf{m}^{*\prime\prime}\times\mathbf{m}''}
\end{aligned}
$$

Secondly

$$
\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m}'}\big[[\mu_{\mathbf{m}}]_{l}\otimes[\psi_{\mathbf{m}\mathbf{m}'}]_{l''\times l'''}\otimes[\mu_{\mathbf{m}'}]_{l'}\big] = [E_{\mathbf{m}}]_{l\times l''\times\mathbf{L}^{**}\mathbf{N}^{**}}\,[E_{\mathbf{m}'}]_{l'\times l'''\times\mathbf{L}^{**}\mathbf{N}^{**}}
$$

where

$$
\begin{aligned}
[E_{\mathbf{m}}]_{l\times l''\times\mathbf{L}^{**}\mathbf{N}^{**}} &:= \Big([K_Y]^{-1/2}_{\mathbf{L}^{**}\mathbf{N}^{**}\times\mathbf{L}^{*\prime\prime}\mathbf{N}^{*\prime\prime}}\otimes[g_{\mathbf{0}}KY]^{\dagger}_{l\times\mathbf{L}^{*}\mathbf{N}^{*}}\Big) \\
&\left[\frac{[g_{\mathbf{0}}]_{l''\times\mathbf{L}^{*\prime\prime}\times\mathbf{N}^{*\prime\prime}}\circ \mathrm{p}\big([\Phi]_{l\times\mathbf{L}^{*}}\,[G]_{\mathbf{m}\times l''\times\mathbf{L}^{*\prime\prime}\times\mathbf{N}^{*\prime\prime}}\big|\,[G]_{\mathbf{m}\times l\times\mathbf{L}^{*}\times\mathbf{N}^{*}}\,,[D]\big)}{\mathrm{p}\big([0]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times\mathbf{L}^{*}\times\mathbf{N}^{*}}\,,[\Phi]_{l\times\mathbf{L}^{*}}\big)}\right]^{\dagger}
\end{aligned}
$$

$$
\begin{aligned}
[D]_{l\times l''\times l^{*}\times l^{*\prime\prime}\times\mathbf{m}^{2}} &:= [\Phi]_{l\times l^{*}\times\mathbf{m}\times\mathbf{m}''} \\
&\quad - [\Phi]_{l\times l^{*}\times\mathbf{m}\times\mathbf{m}^{*\prime\prime}}\,[\Phi]_{l''\times l^{*\prime\prime}\times\mathbf{m}^{*\prime\prime}\times\mathbf{m}^{*}}\,[\Phi]_{l\times l^{*}\times\mathbf{m}^{*}\times\mathbf{m}''}
\end{aligned}
$$

and $[E_{\mathbf{m}'}]_{l'\times l'''\times\mathbf{L}^{**}\mathbf{N}^{**}}$ substitutes $\mathbf{m}\mapsto\mathbf{m}'$, $l\mapsto l'$, $l''\mapsto l'''$ in these definitions. In other words, add a prime superscript to every symbol which is not superscripted **.

This completes the calculation of all quantities of interest.

6.3. Simplifications

The tensors calculated in this Section have many dimensions, up to L^4N^3 each. For a medium-sized problem of 10 outputs and 1000 datapoints, this is 10^{13} dimensions, for larger problems this could easily breach 10^{20} dimensions. This will challenge the memory limitations of a CPU or GPU. However, there are two simplifications which substantially ease this burden.

Firstly, if the signal covariance $[F]_{\mathbf{L}^2}$ is diagonal (in which case the lengthscales tensor $[F]_{\mathbf{L}^2\times\mathbf{m}}$ may as well be diagonal in $\mathbf{L}^2$) then

$$
l''=l\ ;\ l'''=l' \quad \text{throughout Section 6.1.}
$$

$$
l'''=l'' \text{ and } l^{*\prime\prime}=l''\ ;\ l^{*}=l\ ;\ l^{*\prime\prime\prime}=l'''\ ;\ l^{*\prime}=l' \quad \text{throughout Section 6.2.}
$$

This reduces the largest tensor $[H_{\mathbf{m}}]$ needed to calculate a Sobol' index $S_{\mathbf{m}}$ from L^4N^2 dimensions to L^2N^2. The same reduction factor L^2 applies to the two largest terms needed to calculate the variance $T_{\mathbf{m}}$. Note that the noise variance $[E]_{\mathbf{L}^2}$ need not be diagonal to achieve these reductions.

Secondly, to assess uncertainty we are only really interested in the variances of Sobol' indices, not the cross covariances between them. This means

$$l' = l \ ; \ l''' = l'' \quad \text{or} \quad l' = l'' \ ; \ l''' = l \quad \text{throughout Section 6.2.}$$

This reduces the largest tensor in the calculation of $T_{\mathbf{m}}$ from L^4N^3 dimensions to L^2N^3, or from L^2N^3 to LN^3 if the first simplification has already been applied.

7. Conclusion

In this paper, we transformed uniformly distributed inputs u to normally distributed inputs z, and arbitrarily roated them, prior to fitting a Multi-Output Gaussian Process (MOGP). This is restricted to an anisotropic radial basis function (RBF/ARD) kernel, broadly applicable to smoothly varying outputs. Using this surrogate, analytic expressions for closed Sobol' indices $S_{\mathbf{m}}$ are given by Eqs. (3) and (11) and Section 6.1. Analytic expressions for the variance of these estimates over ungovernable noise is given by Eqs. (5), (12) and (13) and Section 6.2. Cheap simplifications of the results are described in Section 6.3.

The use of a surrogate should greatly ease data requirements, preliminary tests indicating that 100 datapoints is sufficient for Sobol' indices within 10% accuracy, as opposed to 10,000+ datapoints for direct calculation. However, we would caution against allowing more than 15 inputs in any case, as all GPs are extremely susceptible to every aspect of the curse of dimensionality Perhaps the most significant feature of this work is the natural cancellation of the regression noise E from the calculation of Sobol' indices, and their variances.

This allows for an arbitrary rotation Θ of inputs, as normal variables are additive, whereas summing uniform inputs does not produce uniform inputs. If the goal is reducing inputs, rotating their basis first boosts the possibilities immensely [22]. It presents the possibilty of choosing Θ to maximise the closed Sobol' index of the first few inputs.

The quantities to be calculated and their formal context are introduced in Section 2. Our approach effectively regards a regression model furnishing an uncertainty measure with each prediction as just another name for a stochastic process. A great deal of progress is made in Section 3 using general stochastic (not necessarily Gaussian) processes. This approach is analytically cleaner, as it is not obfuscated by the GP details. Furthermore, it turns out that the desirable properties of the Gaussian (lack of skew, simple kurtosis) are not actually helpful, as these terms cancel of their own accord. This development leaves just two terms to be calculated, which require the stochastic process to be specified. MOGPs with an RBF/ARD kernel are tersely developed and described in Section 4, then used to calculate the two unknown terms in Sections 5 and 6. Conclusions are drawn in Section 7.

In order to further elucidate these estimates, we must fill in the details of the underlying stochastic processes, sufficiently identifying the regression $\mathtt{y}$ by its first two moments μ_1, μ_2. Then all the answers we desire are given by Eqs. (3) and (11), and Eqs. (5), (12) and (13).

References

[1] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. L. Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, H. R. Maier, The future of sensitivity analysis: An essential discipline for systems modeling and policy support 137 104954. `doi:10.1016/j.envsoft.2020.104954`.

[2] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, Q. Wu, Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices 114 29–39. `doi:10.1016/j.envsoft.2019.01.012`.

[3] R. Ghanem, D. Higdon, H. Owhadi, et al., Handbook of uncertainty quantification, Vol. 6, Springer, 2017.

[4] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280.

[5] D. Chicco, M. J. Warrens, G. Jurman, The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation 7 e623. doi:10.7717/peerj-cs.623.

[6] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, Sensitivity indices for multivariate outputs, Comptes Rendus Mathematique 351 (7-8) (2013) 307–310. doi:10.1016/j.crma.2013.04.016.

[7] B. Lamoureux, N. Mechbal, J. R. Massé, A combined sensitivity analysis and kriging surrogate modeling for early validation of health indicators, Reliability Engineering and System Safety 130 (2014) 12–26. doi:10.1016/j.ress.2014.03.007.

[8] R. G. Ghanem, P. D. Spanos, Spectral techniques for stochastic finite elements, Archives of Computational Methods in Engineering 4 (1) (1997) 63–100. doi:10.1007/BF02818931.

[9] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644. doi:10.1137/s1064827501387826.

[10] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.

[11] M. Chevreuil, R. Lebrun, A. Nouy, P. Rai, A least-squares method for sparse low rank approximation of multivariate functions, SIAM/ASA Journal on Uncertainty Quantification 3 (1) (2015) 897–921. arXiv:http://arxiv.org/abs/1305.0030v2, doi:10.1137/13091899X.

[12] K. Konakli, B. Sudret, Global sensitivity analysis using low-rank tensor approximations, Reliability Engineering & System Safety 156 (2016) 64–83. doi:10.1016/j.ress.2016.07.012.

[13] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[14] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical Science 4 (4) (1989) 409–423.

[15] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series), The MIT Press, 2005.

[16] M. A. Alvarez, L. Rosasco, N. D. Lawrence, Kernels for vector-valued functions: a reviewarXiv:http://arxiv.org/abs/1106.6251v2.

[17] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 751–769. doi:10.1111/j.1467-9868.2004.05304.x.

[18] W. Chen, R. Jin, A. Sudjianto, Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty, Journal of Mechanical Design 127 (5) (2005) 875. doi:10.1115/1.1904642.

[19] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of sobol indices for the gaussian process metamodel, Reliability Engineering & System Safety 94 (3) (2009) 742–751. doi:10.1016/j.ress.2008.07.008.

[20] A. Srivastava, A. K. Subramaniyan, L. Wang, Analytical global sensitivity analysis with gaussian processes, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 31 (03) (2017) 235–250. doi:10.1017/s0890060417000142.

[21] Z. Wu, D. Wang, P. O. N, F. Hu, W. Zhang, Global sensitivity analysis using a gaussian radial basis function metamodel, Reliability Engineering & System Safety 154 (2016) 171–179. doi:10.1016/j.ress.2016.06.006.

[22] P. G. Constantine, Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies, Society for Industrial and Applied Mathematics, 2015. doi:10.1137/1.9781611973860.

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

[24] A. Villani, Another note on the inclusion $\mathrm{L}_p(\mu) \subset \mathrm{L}_q(\mu)$, The American Mathematical Monthly 92 (7) (1985) 485. doi:10.2307/2322503.

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

[26] NumPy user guide: Broadcasting (2022).
URL https://numpy.org/doc/stable/user/basics.broadcasting.html

[27] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy 585 (7825) 357–362. doi:10.1038/s41586-020-2649-2.

[28] 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