Multi-level and multi-fidelity models

Richard Wilkinson

School of Maths and Statistics
University of Sheffield

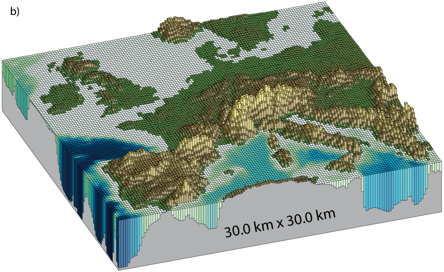
June 2018

Hierarchy of models

High-fidelity model

$$f_{hi} : \mathcal{X} \to \mathcal{Y}$$

Accurate(?) and costly

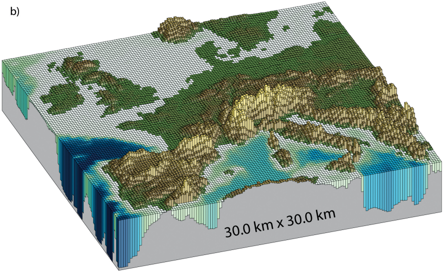


Hierarchy of models

High-fidelity model

$$f_{hi} : \mathcal{X} \to \mathcal{Y}$$

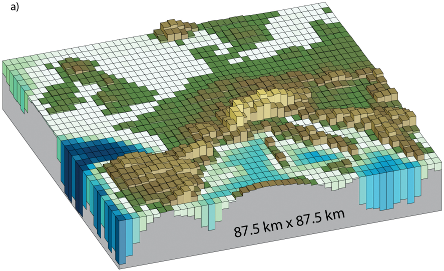
Accurate(?) and costly



Low-fidelity model

$$f_{lo} : \mathcal{X} \to \mathcal{Y}$$

Less accurate and less costly



The low-fidelity models estimate the same quantity from the same inputs, but with lower cost and lower accuracy.

Hierarchy of models

In general, we may have a sequence of models, with $f_{hi} = f^{(1)}$

$$f^{(i)} : \mathcal{X} \to \mathcal{Y} \text{ for } i = 1, \ldots, k$$

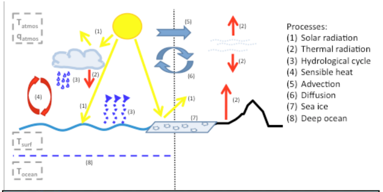
Hierarchy of models

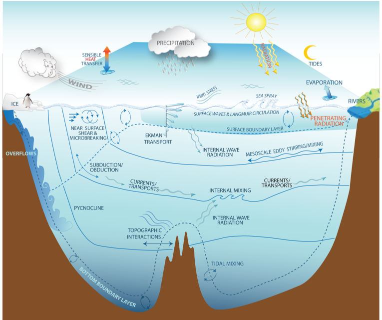
In general, we may have a sequence of models, with $f_{hi} = f^{(1)}$

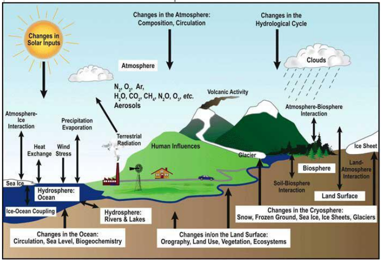
$$f^{(i)} : \mathcal{X} \rightarrow \mathcal{Y} \text{ for } i = 1, \ldots, k$$

These models may (multilevel) or may not (multifidelity) be structurally related

$$\frac{\mathrm{d}X_t}{\mathrm{d}t} = g(X_t, \theta) + F(t, \gamma)$$







Aim of multilevel/multifidelity modelling

- Use the model hierarchy to learn from the high-fidelity model, but at a lower computational cost.

  Equivalently, learn better (lower variance) for the same cost.

Aim of multilevel/multifidelity modelling

- Use the model hierarchy to learn from the high-fidelity model, but at a lower computational cost.

  Equivalently, learn better (lower variance) for the same cost.
- ML/MF modelling is focussed on the best inference for the high-fidelity model - not extracting different types of information from an ensemble of models
  i.e. different to usual aim of having multi-model ensembles.

‘Outer-loop’ applications

Peherstorfer, Willcox, Gunzburger 2017

Analysis that requires multiple calls to f:

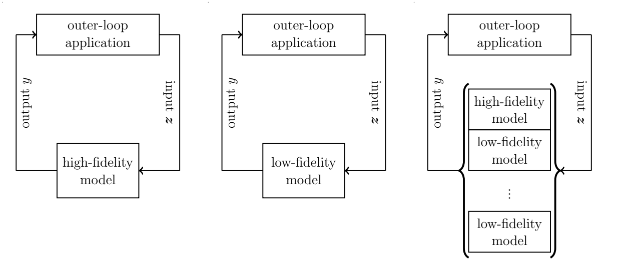
- Uncertainty propagation, eg, $\mathbb{E}(f_{hi}) = \int f_{hi}(x)\pi(x)dx$
- Inference $\pi(x|D) \propto \pi(D|x)\pi(x)$ where $D = f_{hi}(x) + \epsilon$ for some x.
- Optimization, eg, $x_{opt} = \arg\max_x f_{hi}(x)$

'Outer-loop' applications

Peherstorfer, Willcox, Gunzburger 2017

Analysis that requires multiple calls to f:

- Uncertainty propagation, eg, $\mathbb{E}(f_{hi}) = \int f_{hi}(x)\pi(x)dx$
- Inference $\pi(x|D) \propto \pi(D|x)\pi(x)$ where $D = f_{hi}(x) + \epsilon$ for some x.
- Optimization, eg, $x_{opt} = \arg\max_x f_{hi}(x)$

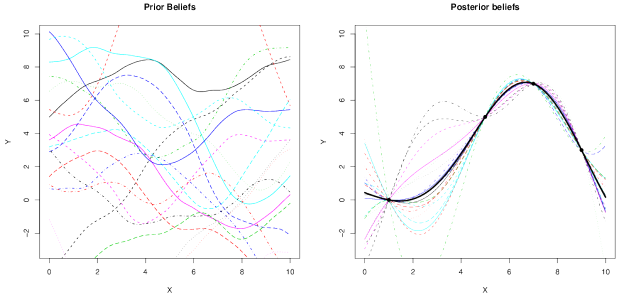


(a) single-fidelity approach with high-fidelity model (b) single-fidelity approach with low-fidelity model (c) multifidelity approach with high-fidelity model and multiple low-fidelity models

Statistical approaches

GPs, GPs, GPs - Hurrah, another nail!

Primarily based around the use of Gaussian processes

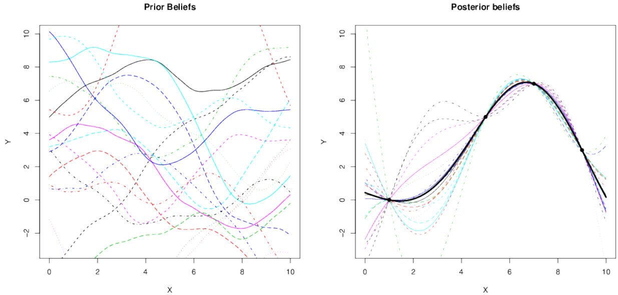


Think of these as probabilistic models of functions.

Statistical approaches

GPs, GPs, GPs - Hurrah, another nail!

Primarily based around the use of Gaussian processes



Think of these as probabilistic models of functions.

A GP is a random process indexed by $x \in \mathcal{X}$ say, such that for every finite set of indices, $x_1, \ldots, x_n$,

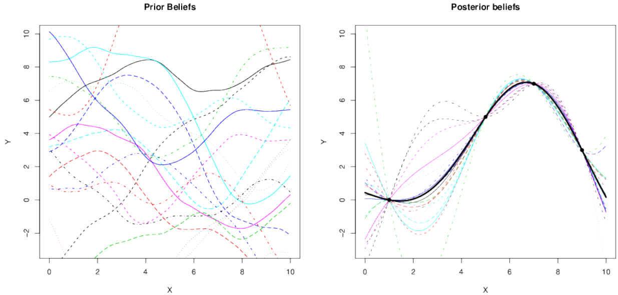
$$\mathbf{f} = (f(x_1), \ldots, f(x_n))$$

has a multivariate Gaussian distribution.

Statistical approaches

GPs, GPs, GPs - Hurrah, another nail!

Primarily based around the use of Gaussian processes



Think of these as probabilistic models of functions.

A GP is a random process indexed by $x \in \mathcal{X}$ say, such that for every finite set of indices, $x_1, \ldots, x_n$,

$$\mathbf{f} = (f(x_1), \ldots, f(x_n))$$

has a multivariate Gaussian distribution.

Why would we want to use this very restricted model?

Answer 1

Class of models is closed under various operations.

Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

- Closed under Bayesian conditioning, i.e., if we observe

$$\mathbf{D} = (f(x_1), \ldots, f(x_n))$$

then

$$f|D \sim GP$$

but with updated mean and covariance functions.

Answer 1

Class of models is closed under various operations.

- Closed under addition

$$f_1(\cdot), f_2(\cdot) \sim GP \quad \text{then} \quad (f_1 + f_2)(\cdot) \sim GP$$

- Closed under Bayesian conditioning, i.e., if we observe

$$\mathbf{D} = (f(x_1), \ldots, f(x_n))$$

  then

$$f|D \sim GP$$

  but with updated mean and covariance functions.

- Closed under any linear operation. If $\mathcal{L}$ is a linear operator, then

$$\mathcal{L}f \sim GP(\mathcal{L}m, \mathcal{L}k\mathcal{L}^\top)$$

  e.g. $\frac{df}{dx}$, $\int f(x)dx$, Af are all GPs

Answer 2: non-parametric/kernel regression

- Linear regression $y = x^\top \beta + \epsilon$ can be written solely in terms of inner products $x^\top x$.

$$\begin{aligned}\hat{\beta} &= \arg\min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2 \\ &= X^\top (XX^\top + \sigma^2 I)^{-1} y \quad \text{(the dual form)}\end{aligned}$$

Answer 2: non-parametric/kernel regression

- Linear regression $y = x^\top \beta + \epsilon$ can be written solely in terms of inner products $x^\top x$.

$$\begin{aligned}\hat{\beta} &= \arg\min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2 \\ &= X^\top (XX^\top + \sigma^2 I)^{-1} y \quad \text{(the dual form)}\end{aligned}$$

- We know that we can replace x by a feature vector in linear regression, e.g., $\phi(x) = (1 \; x \; x^2 \; \cos(x))^\top$ etc.

Answer 2: non-parametric/kernel regression

- Linear regression $y = x^\top \beta + \epsilon$ can be written solely in terms of inner products $x^\top x$.

$$\hat{\beta} = \arg\min ||y - X\beta||_2^2 + \sigma^2 ||\beta||_2^2$$
$$= X^\top (XX^\top + \sigma^2 I)^{-1} y \quad \text{(the dual form)}$$

- We know that we can replace x by a feature vector in linear regression, e.g., $\phi(x) = (1\ x\ x^2\ \cos(x))^\top$ etc.
- For some features, inner product is equivalent to evaluating a kernel

$$\phi(x)^\top \phi(x') \equiv k(x, x')$$

where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a semi-positive definite function.

Answer 2: non-parametric/kernel regression

- Linear regression $y = x^\top\beta + \epsilon$ can be written solely in terms of inner products $x^\top x$.

$$\begin{aligned}\hat{\beta} &= \arg\min ||y - X\beta||_2^2 + \sigma^2||\beta||_2^2 \\ &= X^\top(XX^\top + \sigma^2 I)^{-1}y \quad \text{(the dual form)}\end{aligned}$$

- We know that we can replace x by a feature vector in linear regression, e.g., $\phi(x) = (1\ x\ x^2\ \cos(x))^\top$ etc.
- For some features, inner product is equivalent to evaluating a kernel

$$\phi(x)^\top\phi(x') \equiv k(x, x')$$

where $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is a semi-positive definite function.
Kernel trick: lift x into infinite dimensional feature space by replacing inner products $x^\top x'$ by $k(x, x')$, but never evaluate these features, only the $n \times n$ kernel matrix.

$$\hat{y}' = m(x') = \sum_{i=1}^{n} \alpha_i k(x, x_i)$$

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of k).

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of k).

Example: If (modulo some detail)

$$\phi(x) = (e^{-\frac{(x-c_1)^2}{2\lambda^2}}, \ldots, e^{-\frac{(x-c_N)^2}{2\lambda^2}})$$

then as $N \to \infty$ then

$$\phi(x)^\top \phi(x) = \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$$

Generally, we don't think about features, we just choose a kernel. But choosing a kernel is implicitly choosing features, and our model only includes functions that are linear combinations of this set of features (the Reproducing Kernel Hilbert Space (RKHS) of k).

Example: If (modulo some detail)

$$\phi(x) = (e^{-\frac{(x-c_1)^2}{2\lambda^2}}, \ldots, e^{-\frac{(x-c_N)^2}{2\lambda^2}})$$

then as $N \to \infty$ then

$$\phi(x)^\top \phi(x) = \exp\left(-\frac{(x-x')^2}{2\lambda^2}\right)$$

Although our simulator may not lie in the RKHS defined by k, this space is much richer than any parametric regression model (and can be dense in some sets of continuous bounded functions), and is thus more likely to contain an element close to the simulator than any class of models that contains only a finite number of features.

Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

1

Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

One answer might come from Bayes linear methods[1].
If we only knew the expectation and variance of some random variables, X and Y, then how should we best do statistics?

[1] statistics without probability

Answer 3: Naturalness of GP framework

Why use **Gaussian** processes as non-parametric models?

One answer might come from Bayes linear methods[1].
If we only knew the expectation and variance of some random variables, X and Y, then how should we best do statistics?

It can been shown, that the best second-order inference we can do to update our beliefs about X given Y is

$$\mathbb{E}(X|Y) = \mathbb{E}(X) + \mathbb{C}\text{ov}(X, Y)\mathbb{V}\text{ar}(Y)^{-1}(Y - \mathbb{E}(Y))$$

which is exactly the Gaussian process update for the posterior mean.

So GPs are in some sense very natural approaches.

[1] statistics without probability

Multifidelity modelling via co-kriging

Craig *et al.* 1998, Kennedy *et al.* 2000, Cumming *et al.* 2009

Using a GP to approximate f_{hi} is already a multi-fidelity approach in the language of Peherstorfer *et al.*

In statistics UQ, multilevel usually refers to approaches where we model multiple mechanistic models, $f^{(1)}, \ldots, f^{(k)}$, each with a Gaussian process, linking the emulators at each level.

Multifidelity modelling via co-kriging

Craig *et al.* 1998, Kennedy *et al.* 2000, Cumming *et al.* 2009

Using a GP to approximate f_{hi} is already a multi-fidelity approach in the language of Peherstorfer *et al.*

In statistics UQ, multilevel usually refers to approaches where we model multiple mechanistic models, $f^{(1)}, \ldots, f^{(k)}$, each with a Gaussian process, linking the emulators at each level.

We build surrogate models at each level of the model hierarchy, ie, given training sets

$$D_i = \{(x_j, y_j = f^{(i)}(x_i))_{j=1,\ldots,n}\}$$

build emulator $F^{(i)}$ of $f^{(i)}$, linking the emulators at each level.

Co-kriging

Kennedy and O'Hagan 2000 etc

If Markov property holds: $F^{(i)}(x) \perp\!\!\!\perp F^{(i+1)}(x') | F^{(i+1)}(x)$ then

$$F^{(i)}(x) = \rho_i F^{(i+1)}(x) + \delta_i(x)$$
$$F^{(i+1)}(x) \perp\!\!\!\perp \delta_i(x)$$

where

$$\delta_i(\cdot), Z^{(i)}(\cdot) \sim GP(m(\cdot), c(\cdot, \cdot))$$

Le Gratiet and Garnier 2014 improve upon this by setting

$$F^{(i)}(x) = \rho_i(x) \hat{F}^{(i+1)}(x) + \delta_i(x)$$

where $\hat{F}^{(i+1)}$ is a GP with distribution $[F^{(i+1)}(x) | D_{i+1}]$
This doesn't change the prediction but is computationally cheaper, has closed form expressions for the mean and variances, and has easily computable cross-validation approximations.

Statistical approaches

Perdikaris *et al.* 2017 go further with the modelling:

$$F^{(i)}(x) = g_i(F^{(i-1)}(x)) + \delta_i(x)$$

where $g_i(\cdot)$ is another GP, resulting in a *deep* GP structure.

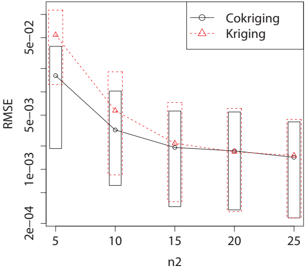
Statistical approaches

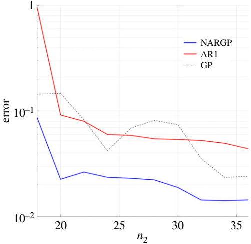
Perdikaris *et al.* 2017 go further with the modelling:

$$F^{(i)}(x) = g_i(F^{(i-1)}(x)) + \delta_i(x)$$

where $g_i(\cdot)$ is another GP, resulting in a *deep* GP structure.
Once the emulator is built, it is used directly in the outer-loop application. Generally, the approach is shown to work via demonstration on a few examples.

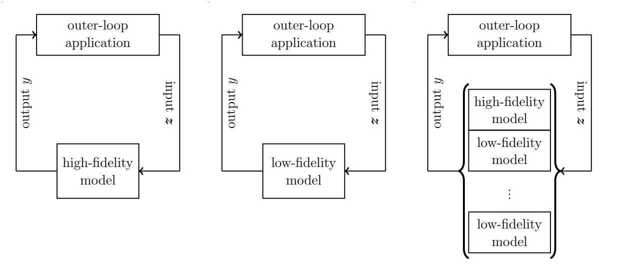
- no guarantees on performance
- typical problems allow 10s of high fidelity runs.
- methods suitable for black-box models





Applied math ML/MF

Peherstorfer, Willcox, Gunzburger 2017



(a) single-fidelity approach with high-fidelity model (b) single-fidelity approach with low-fidelity model (c) multifidelity approach with high-fidelity model and multiple low-fidelity models

Focus on methods which have guarantees on the outer-loop result.

- Typically methods determine a *model management strategy*, which distributes work amongst the models whilst providing theoretical guarantes establishing accuracy and/or convergence of outer-loop result.

Example: Multi-fidelity Uncertainty Propagation

Control variates

Basic idea:

- m an unbiased estimator of μ so that $\mathbb{E}(m) = \mu$
- t a random variable with $\mathbb{E}(t) = \tau$
- Then

$$m^* = m + c(t - \tau)$$

  is also an unbiased estimator of μ for any c.

Example: Multi-fidelity Uncertainty Propagation

Control variates

Basic idea:

- m an unbiased estimator of μ so that $\mathbb{E}(m) = \mu$
- t a random variable with $\mathbb{E}(t) = \tau$
- Then

  $$m^* = m + c(t - \tau)$$

  is also an unbiased estimator of μ for any c.
- The optimal choice is $c = -\mathbb{C}\text{ov}(m, t)/\mathbb{V}\text{ar}(t)$ and then

  $$\mathbb{V}\text{ar}(m^*) = (1 - \rho^2)\mathbb{V}\text{ar}(m)$$

  where $\rho = \text{corr}(m, t)$

So if we can find an estimator t that is highly correlated with m we can greatly improve our estimator.

Example: Multi-fidelity Uncertainty Propagation

Peherstorfer, Willcox, Gunzburger 2016

Target: $s = \mathbb{E} f^{(1)}(X)$

Example: Multi-fidelity Uncertainty Propagation

Peherstorfer, Willcox, Gunzburger 2016

$$\text{Target: } s = \mathbb{E} f^{(1)}(X)$$
$$\approx \frac{1}{m} \sum_{j=1}^{m} f^{(1)}(x_j) := \bar{y}_m^{(1)}$$

Example: Multi-fidelity Uncertainty Propagation

Peherstorfer, Willcox, Gunzburger 2016

$$\text{Target: } s = \mathbb{E} f^{(1)}(X) \approx \frac{1}{m} \sum_{j=1}^{m} f^{(1)}(x_j) := \bar{y}_m^{(1)}$$

We create control variates by nesting the evaluation of the low-fidelity simulators.

Example: Multi-fidelity Uncertainty Propagation

Peherstorfer, Willcox, Gunzburger 2016

$$\text{Target: } s = \mathbb{E} f^{(1)}(X)$$
$$\approx \frac{1}{m} \sum_{j=1}^{m} f^{(1)}(x_j) := \bar{y}_m^{(1)}$$

We create control variates by nesting the evaluation of the low-fidelity simulators.

We'll do m_1 evaluations of $f^{(1)}$, m_2 evaluations of $f^{(2)}$ etc with $m_i < m_{i+1}$
Given random samples $X_1, \ldots, X_{m_1}, \ldots, X_{m_2}, \ldots, X_{m_k}$ form estimator

$$\hat{s} = \bar{y}_{m_1}^{(1)} + \sum_{i=2}^{k} \alpha_i (\bar{y}_{m_i}^{(i)} - \bar{y}_{m_{i-1}}^{(i)})$$

Example: Multi-fidelity Uncertainty Propagation

Peherstorfer, Willcox, Gunzburger 2016

$$\begin{aligned}\text{Target: } s &= \mathbb{E} f^{(1)}(X) \\ &\approx \frac{1}{m}\sum_{j=1}^{m} f^{(1)}(x_j) := \bar{y}_m^{(1)}\end{aligned}$$

We create control variates by nesting the evaluation of the low-fidelity simulators.

We'll do m_1 evaluations of $f^{(1)}$, m_2 evaluations of $f^{(2)}$ etc with $m_i < m_{i+1}$

Given random samples $X_1, \ldots, X_{m_1}, \ldots, X_{m_2}, \ldots, X_{m_k}$ form estimator

$$\hat{s} = \bar{y}_{m_1}^{(1)} + \sum_{i=2}^{k} \alpha_i (\bar{y}_{m_i}^{(i)} - \bar{y}_{m_{i-1}}^{(i)})$$

$\hat{s}$ is obviously unbiased for s.

Peherstorfer *et al.* solve the optimization problem

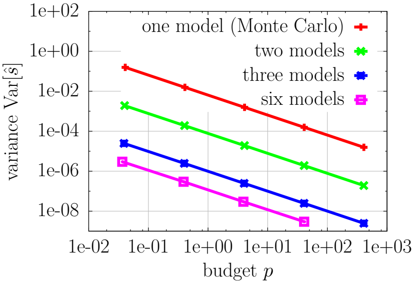
$$\begin{aligned} \min_{\mathbf{m}\in\mathbb{R}^k,\alpha_2,\ldots,\alpha_k\in\mathbb{R}} & \mathbb{V}\text{ar}(\hat{s}) \\ \text{s.t. } m_1 &> 0 \\ m_i &> m_{i-1} \\ \mathbf{m}^\top \mathbf{c} &= \text{budget} \end{aligned}$$

for given simulator costs $c_1, \ldots, c_k$.
The solution is a function of correlations $\rho_{1,i} = \text{cor}(f^{(1)}(X), f^{(i)}(X))$.

Peherstorfer *et al.* solve the optimization problem

$$
\begin{aligned}
\min_{\mathbf{m}\in\mathbb{R}^k,\alpha_2,\ldots,\alpha_k\in\mathbb{R}} \quad & \mathbb{V}\mathrm{ar}(\hat{s}) \\
\text{s.t.} \quad & m_1 > 0 \\
& m_i > m_{i-1} \\
& \mathbf{m}^\top \mathbf{c} = \text{budget}
\end{aligned}
$$

for given simulator costs $c_1, \ldots, c_k$.
The solution is a function of correlations $\rho_{1,i} = \mathrm{cor}(f^{(1)}(X), f^{(i)}(X))$.



- Plate bending problem with $c_1/c_2 = 10^2$ and $\rho_{1,2} = 0.99999$
- Note there are no assumptions on the surrogate, i.e., no bounds on

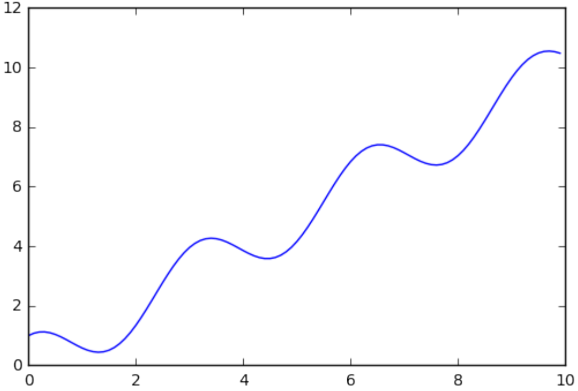
$$|f^{(1)}(x) - f^{(i)}(x)|$$

- Only require the correlations $\rho_{1,i}$

Combining multifidelity MC with GP emulation

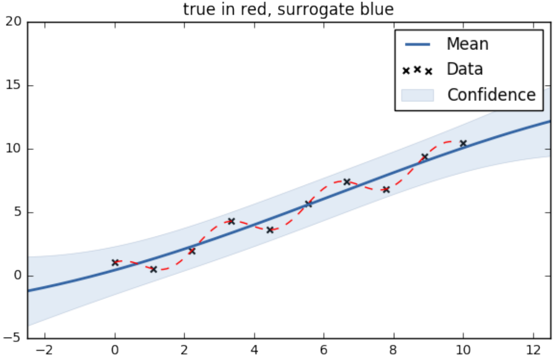
Imagine we have a expensive function f for which we want to estimate

$$\mathbb{E}f(X) = \int_0^{10} \frac{f(x)}{10} dx$$



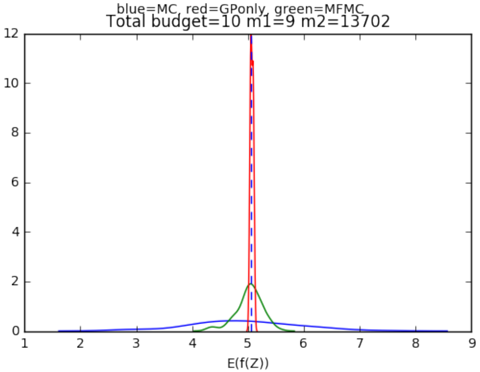
Combining multifidelity MC with GP emulation

Build a GP emulator:



Combining multifidelity MC with GP emulation

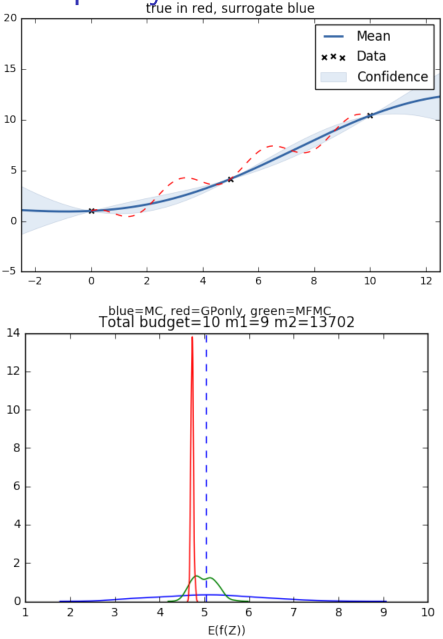
Use i) Monte Carlo, ii) just the GP, and iii) multifidelity Monte Carlo to estimate the expectation. Repeat the procedure to get an idea of sampling variation.



Total budget here is 10 expensive simulator evaluations, and I've assumed

$$\frac{c_1}{c_2} = 10^5$$

Lower quality emulator



For a good emulator, the MFMC estimate is worse than the estimate which just naively uses the GP.

However, the uncertainty estimates for GP emulators are often poor, particularly for high dimensional problems.

For a poor emulator, MFMC unbiases the estimate.

Problems of using GPs with MFMC

- The method requires $\sigma_i^2 = \mathbb{V}\mathrm{ar} f^{(i)}(X)$ and $\rho_{1,i}$.
  - Estimating these is harder than estimating $s = \mathbb{E}(f^{(1)}(x))$
  - Do poor estimates reduce or eliminate the benefit of MFMC?

Problems of using GPs with MFMC

- The method requires $\sigma_i^2 = \mathbb{V}\text{ar} f^{(i)}(X)$ and $\rho_{1,i}$.
  - Estimating these is harder than estimating $s = \mathbb{E}(f^{(1)}(x))$
  - Do poor estimates reduce or eliminate the benefit of MFMC?
- When using GP emulators, for MFMC we'd need two or three training sets
  - train the emulator
  - estimate the correlations and variances
  - form the MFMC estimator

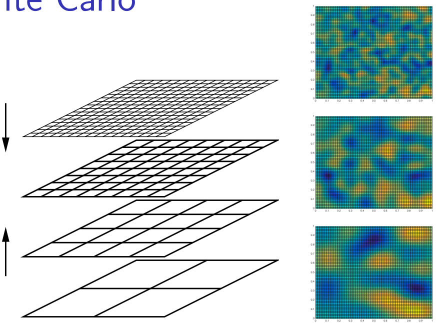
  We can't directly use the emulator training set to estimate correlations.

Problems of using GPs with MFMC

- The method requires $\sigma_i^2 = \mathbb{V}\mathrm{ar} f^{(i)}(X)$ and $\rho_{1,i}$.
  - Estimating these is harder than estimating $s = \mathbb{E}(f^{(1)}(x))$
  - Do poor estimates reduce or eliminate the benefit of MFMC?
- When using GP emulators, for MFMC we'd need two or three training sets
  - train the emulator
  - estimate the correlations and variances
  - form the MFMC estimator

  We can't directly use the emulator training set to estimate correlations.
  - Can bootstrapping approaches reduce the number of simulator evaluations necessary and give a MFMC-GP approach which is guaranteed to be unbiased?

Multi-level Monte Carlo



Multi-level Monte Carlo predates MFMC (eg Giles 2008)

- Assumes we have models that result from the coarsening of a fine computational mesh.
- The error rates and costs for each level are typically known theoretically for each type of problem
- Similar kind of nesting idea (again exploiting control variates idea) used to form an estimator

Inference

Grey box models: physically obedient GPs

Black box methods use no knowledge of the underlying equations in the model
Intrusive methods require complete knowledge

Grey box models: physically obedient GPs

Black box methods use no knowledge of the underlying equations in the model
Intrusive methods require complete knowledge
Can we develop 'grey-box' methods?
E.g. suppose model output is $f(x)$ where f is the solution of

$$\mathcal{F}_x^1[f] = 0$$
$$\mathcal{F}_x^2[f] = w(x)$$
$$\vdots$$

Can we find GP emulators that obey simpler constraints exactly, and use data to train to the other constraints?

Grey box models: physically obedient GPs

Black box methods use no knowledge of the underlying equations in the model
Intrusive methods require complete knowledge
Can we develop 'grey-box' methods?
E.g. suppose model output is $f(x)$ where f is the solution of

$$\mathcal{F}_x^1[f] = 0$$
$$\mathcal{F}_x^2[f] = w(x)$$
$$\vdots$$

Can we find GP emulators that obey simpler constraints exactly, and use data to train to the other constraints?
E.g., guarantee that $\nabla . f = 0$ or $\nabla \times f = 0$ etc.

Grey box models: physically obedient GPs

Jidling *et al.* 2017

Simple idea: Suppose $f = \mathcal{G}_x[g]$ for some linear operator $\mathcal{G}_x$ so that for any function g, f satisfies $\mathcal{F}_x[f] = 0$ for linear operator $\mathcal{F}_x$.

Grey box models: physically obedient GPs

Jidling *et al.* 2017

Simple idea: Suppose $f = \mathcal{G}_x[g]$ for some linear operator $\mathcal{G}_x$ so that for any function g, f satisfies $\mathcal{F}_x[f] = 0$ for linear operator $\mathcal{F}_x$.
e.g. if $f : \mathbb{R}^2 \to \mathbb{R}^2$ and

$$\mathcal{F}_x = \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{pmatrix} \qquad \text{ie} \quad \mathcal{F}_x[f] = \nabla . f$$

Grey box models: physically obedient GPs

Jidling *et al.* 2017

Simple idea: Suppose $f = \mathcal{G}_x[g]$ for some linear operator $\mathcal{G}_x$ so that for any function g, f satisfies $\mathcal{F}_x[f] = 0$ for linear operator $\mathcal{F}_x$.
e.g. if $f : \mathbb{R}^2 \to \mathbb{R}^2$ and

$$\mathcal{F}_x = \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{pmatrix} \qquad \text{ie} \quad \mathcal{F}_x[f] = \nabla . f$$

then if

$$\mathcal{G}_x = \begin{pmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{pmatrix}$$

we have $f = \mathcal{G}_x[g]$ satisfies $\mathcal{F}_x f = 0$ for all functions $g : \mathbb{R}^2 \to \mathbb{R}$.

Grey box models: physically obedient GPs

Jidling *et al.* 2017

Simple idea: Suppose $f = \mathcal{G}_x[g]$ for some linear operator $\mathcal{G}_x$ so that for any function g, f satisfies $\mathcal{F}_x[f] = 0$ for linear operator $\mathcal{F}_x$.
e.g. if $f : \mathbb{R}^2 \to \mathbb{R}^2$ and

$$\mathcal{F}_x = \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{pmatrix} \qquad \text{ie} \quad \mathcal{F}_x[f] = \nabla . f$$

then if

$$\mathcal{G}_x = \begin{pmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{pmatrix}$$

we have $f = \mathcal{G}_x[g]$ satisfies $\mathcal{F}_x f = 0$ for all functions $g : \mathbb{R}^2 \to \mathbb{R}$.
If $g \sim GP(m(\cdot), k(\cdot, \cdot))$ then

$$f = \mathcal{G}_x[g] \sim GP(\mathcal{G}_x[m], \mathcal{G}_x k \mathcal{G}_x'^{\top})$$

So we can train emulators of f that satisfy part of the model equations.

Grey box models: physically obedient GPs

Jidling *et al.* 2017

Simple idea: Suppose $f = \mathcal{G}_x[g]$ for some linear operator $\mathcal{G}_x$ so that for any function g, f satisfies $\mathcal{F}_x[f] = 0$ for linear operator $\mathcal{F}_x$.
e.g. if $f : \mathbb{R}^2 \to \mathbb{R}^2$ and

$$\mathcal{F}_x = \begin{pmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} \end{pmatrix} \qquad \text{ie} \quad \mathcal{F}_x[f] = \nabla . f$$

then if

$$\mathcal{G}_x = \begin{pmatrix} -\frac{\partial}{\partial y} \\ \frac{\partial}{\partial x} \end{pmatrix}$$

we have $f = \mathcal{G}_x[g]$ satisfies $\mathcal{F}_x f = 0$ for all functions $g : \mathbb{R}^2 \to \mathbb{R}$.
If $g \sim GP(m(\cdot), k(\cdot, \cdot))$ then

$$f = \mathcal{G}_x[g] \sim GP(\mathcal{G}_x[m], \mathcal{G}_x k \mathcal{G}_x'^{\top})$$

So we can train emulators of f that satisfy part of the model equations.
To find $\mathcal{G}_x$ such that $\mathcal{F}_x \mathcal{G}_x$ we look for the null space of the operator $\mathcal{F}_x$....

Conclusions

A good many times I have been present at gatherings of [highly-educated] people... who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold...Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's? C.P. Snow, 'The Two Cultures'

Conclusions

> *A good many times I have been present at gatherings of [highly-educated] people... who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold...Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's? C.P. Snow, 'The Two Cultures'*

Equivalent for this workshop?

- What about the real world?
- Where is the theory? Error guarantees?

Conclusions

> *A good many times I have been present at gatherings of [highly-educated] people... who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold...Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's? C.P. Snow, 'The Two Cultures'*

Equivalent for this workshop?

- What about the real world?
  - Stats aim is usually inference about the world, mathematical rigorous approaches are necessarily limited to 'model-land'.
- Where is the theory? Error guarantees?
  - theoretical error bounds essential for all methods in the math community. Stats UQ comes with few guarantees of this kind - and is generally validated empirically (perhaps it is unwise to be scared of mice when there are tigers abroad).

Conclusions

> *A good many times I have been present at gatherings of [highly-educated] people... who have with considerable gusto been expressing their incredulity at the illiteracy of scientists. Once or twice I have been provoked and have asked the company how many of them could describe the Second Law of Thermodynamics. The response was cold...Yet I was asking something which is the scientific equivalent of: Have you read a work of Shakespeare's? C.P. Snow, 'The Two Cultures'*

Equivalent for this workshop?

- What about the real world?
  - Stats aim is usually inference about the world, mathematical rigorous approaches are necessarily limited to 'model-land'.
- Where is the theory? Error guarantees?
  - theoretical error bounds essential for all methods in the math community. Stats UQ comes with few guarantees of this kind - and is generally validated empirically (perhaps it is unwise to be scared of mice when there are tigers abroad).

Thank you for listening!