Multilevel Bayesian Quadrature

François-Xavier Briol
University College London & The Alan Turing Institute

Workshop on “Fusing Simulations with Data Science”,
University of Warwick, Coventry, UK.

Collaborators

Kaiyu Li
(UCL)

Toni Karvonen
(Helsinki)

Dan Giles
(UCL)

Serge Guillas
(UCL)

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.

Intractable Integrals and Expensive Simulations

Setting: Intractable Integrals

- Let $f : \mathcal{X} \to \mathbb{R}$ ($\mathcal{X} \subseteq \mathbb{R}^d$). Want to estimate the intractable integral:

$$\Pi[f] := \int_{\mathcal{X}} f(x)\Pi(dx).$$

- Statisticians' Favourite Solution: use Monte Carlo:

$$\hat{\Pi}_{\mathrm{MC}}[f] := \frac{1}{n}\sum_{i=1}^{n} f(x_i), \qquad \{x_i\}_{i=1}^{n} \sim \Pi.$$

- Evaluating integrands/sampling can be very very expensive!!

Setting: Intractable Integrals

- Let $f : \mathcal{X} \to \mathbb{R}$ ($\mathcal{X} \subseteq \mathbb{R}^d$). Want to estimate the intractable integral:

$$\Pi[f] := \int_{\mathcal{X}} f(x)\Pi(dx).$$

- Statisticians' Favourite Solution: use Monte Carlo:

$$\hat{\Pi}_{\mathrm{MC}}[f] := \frac{1}{n}\sum_{i=1}^{n} f(x_i), \qquad \{x_i\}_{i=1}^{n} \sim \Pi.$$

- Evaluating integrands/sampling can be very very expensive!!

Setting: Intractable Integrals

- Let $f : \mathcal{X} \to \mathbb{R}$ ($\mathcal{X} \subseteq \mathbb{R}^d$). Want to estimate the intractable integral:

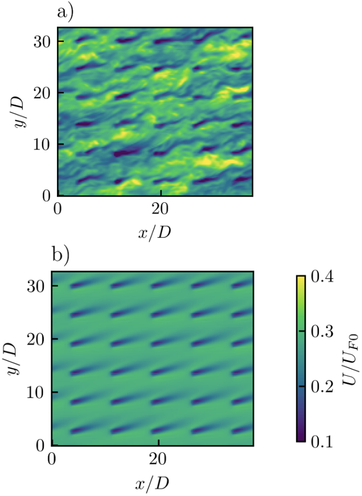
$$\Pi[f] := \int_{\mathcal{X}} f(x)\Pi(dx).$$

- Statisticians' Favourite Solution: use Monte Carlo:

$$\hat{\Pi}_{\text{MC}}[f] := \frac{1}{n}\sum_{i=1}^{n} f(x_i), \qquad \{x_i\}_{i=1}^{n} \sim \Pi.$$

- Evaluating integrands/sampling can be very very expensive!!

Example 1: CFDs for Wind Farm Modelling



Expected energy production of a wind farm given wind conditions.

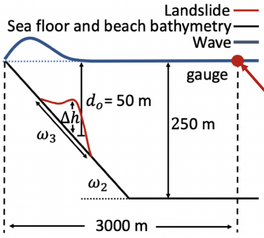
≈ 100 CPU hours per wind farm simulation!!

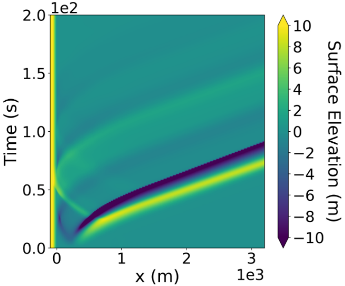
Cannot obtain many samples even with large computational budget.

Kirby, A., Briol, F.-X., Dunstan, T. D., & Nishino, T. (2023). Data-driven modelling of turbine wake interactions and flow resistance in large wind farms. Wind Energy, 1–17.

See Andrew's talk!

Example 2: Uncertainty Quantification for Tsunami Models





PDE-based tsunami model (non-linear shallow water equations). Need to incorporate uncertainty over parameters.

$\approx$ 3 minutes GPU time per integrand evaluations for a simplistic model!

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.

Challenges for Integration in an Expensive World

- We have several key challenges for expensive problems:
  - We will necessarily have a small number of observations n.
  - We will necessarily have significant numerical error remaining after running our integration algorithm.
- We **need** methods which can make use of the structure of the integrand to reduce computational needs!

Challenges for Integration in an Expensive World

- We have several key challenges for expensive problems:
  - We will necessarily have a small number of observations n.
  - We will necessarily have significant numerical error remaining after running our integration algorithm.
- We **need** methods which can make use of the structure of the integrand to reduce computational needs!

Multilevel and Multifidelity Methods

The Telescoping Sum

- **Multifidelity/Multilevel models:** $f_0, \ldots, f_{L-1}$ are approximations of $f := f_L$ of increasing accuracy, but also increasing computational cost.

- The integral can be re-written as:

$$\begin{aligned}\Pi[f] := \Pi[f_L] = & \Pi[f_L - f_{L-1}] \\ & + \Pi[f_{L-1} - f_{L-2}] \\ & + \ldots \\ & + \Pi[f_1 - f_0] \\ & + \Pi[f_0]\end{aligned}$$

The Telescoping Sum

- **Multifidelity/Multilevel models:** $f_0, \ldots, f_{L-1}$ are approximations of $f := f_L$ of increasing accuracy, but also increasing computational cost.

- The integral can be re-written as:

$$\begin{aligned}\Pi[f] := \Pi[f_L] = \; & \Pi[f_L - f_{L-1}] \\ & + \Pi[f_{L-1} - f_{L-2}] \\ & + \ldots \\ & + \Pi[f_1 - f_0] \\ & + \Pi[f_0]\end{aligned}$$

The Telescoping Sum

- **Multifidelity/Multilevel models:** $f_0, \ldots, f_{L-1}$ are approximations of $f := f_L$ of increasing accuracy, but also increasing computational cost.

- The integral can be re-written as:

$$\begin{aligned}\Pi[f] := \Pi[f_L] = & \ \Pi[f_L - f_{L-1}] \\ & + \Pi[f_{L-1} - f_{L-2}] \\ & + \ldots \\ & + \Pi[f_1 - f_0] \\ & + \Pi[f_0]\end{aligned}$$

Multilevel Monte Carlo

- The multilevel Monte Carlo algorithm consists of approximating each term in the sum with a Monte Carlo estimator.

$$\hat{\Pi}_{\mathrm{MLMC}}[f] = \hat{\Pi}_{\mathrm{MC}}[f_L - f_{L-1}] + \hat{\Pi}_{\mathrm{MC}}[f_{L-1} - f_{L-2}] + \dots + \hat{\Pi}_{\mathrm{MC}}[f_1 - f_0] + \hat{\Pi}_{\mathrm{MC}}[f_0]$$

- The idea is to evaluate less often the expensive (but accurate) levels and more often the cheap (but inaccurate) levels.

Giles, M. B. (2015). Multilevel Monte Carlo methods. Acta Numerica, 24, 259–328.

Optimising the performance of MLMC

- Denote by $n^{\text{MLMC}} = (n_0^{\text{MLMC}}, \ldots, n_L^{\text{MLMC}})$ the optimal number of samples per level to obtain a fixed mean squared error. Then:

$$n_l^{\text{MLMC}} \propto \left(\frac{V_l}{C_l}\right)^{\frac{1}{2}} \qquad \text{where} \qquad \begin{cases} V_l = \text{Var}[f_l - f_{l-1}] \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

- This is nice, but MC doesn't use any structure of the integrand.
- Of course there have been a *huge* range of extensions of MLMC:
  - based on points (e.g. importance sampling, QMC, MCMC),
  - clever selection of levels (e.g. multi-index, adaptive selection),
  - specialised algorithms for specific classes of integrands (e.g. based on certain PDEs), etc...

Optimising the performance of MLMC

- Denote by $n^{\text{MLMC}} = (n_0^{\text{MLMC}}, \dots, n_L^{\text{MLMC}})$ the optimal number of samples per level to obtain a fixed mean squared error. Then:

$$n_l^{\text{MLMC}} \propto \left(\frac{V_l}{C_l}\right)^{\frac{1}{2}} \qquad \text{where} \qquad \begin{cases} V_l = \text{Var}[f_l - f_{l-1}] \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

- This is nice, but MC doesn't use any structure of the integrand.
- Of course there have been a *huge* range of extensions of MLMC:
  - based on points (e.g. importance sampling, QMC, MCMC),
  - clever selection of levels (e.g. multi-index, adaptive selection),
  - specialised algorithms for specific classes of integrands (e.g. based on certain PDEs), etc...

Multilevel Bayesian Quadrature

Bayesian Quadrature per level...

A very simple idea:

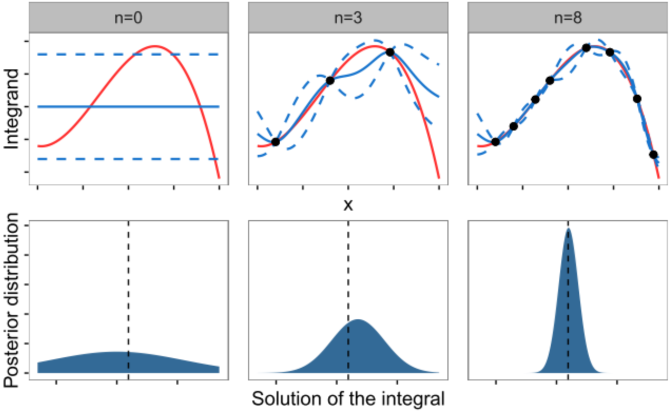
$$\hat{\Pi}_{\text{MLBQ}}[f] = \hat{\Pi}_{\text{BQ}}[f_L - f_{L-1}] + \hat{\Pi}_{\text{BQ}}[f_{L-1} - f_{L-2}] + \dots + \hat{\Pi}_{\text{BQ}}[f_1 - f_0] + \hat{\Pi}_{\text{BQ}}[f_0]$$

Diaconis, P. (1988). Bayesian Numerical Analysis. Statistical Decision Theory and Related Topics IV, 163—175.

O'Hagan, A. (1991). Bayes-Hermite quadrature. Journal of Statistical Planning and Inference, 29, 245-260

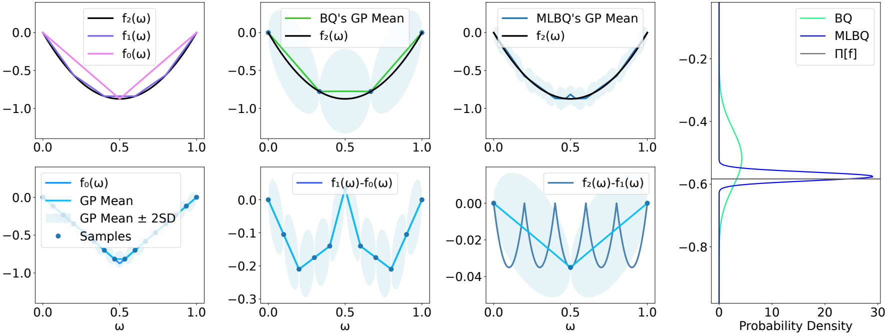
Briol, F.-X., Oates, C. J., Girolami, M., Osborne, M. A., & Sejdinovic, D. (2019). Probabilistic integration: A role in statistical computation? (with discussion). Statistical Science, 34(1), 1-22.

Bayesian Quadrature



1. Posit a Gaussian process (GP) prior distribution on f.
2. Condition this GP prior on data (i.e. $f(X) = (f(x_1), \ldots, f(x_n))^\top$).
3. Consider the pushforward of this distribution on $\Pi[f]$.

Multilevel Bayesian Quadrature



The posterior much more concentrated on the truth!

Multilevel Bayesian Quadrature

- Define $f_{-1} = 0$ for simplicity. Assume $GP(m_l, c_l)$ priors on $f_l - f_{l-1}$ and independence across levels, then, given observations

$$f_l(X_l) - f_{l-1}(X_l) = (f_l(x_{l1}) - f_{l-1}(x_{l1}), \ldots, f_l(x_{ln_l}) - f_{l-1}(x_{ln_l}))^\top$$

for all l, the posterior mean is:

$$\hat{\Pi}_{\mathrm{MLBQ}}[f] := \sum_{l=0}^{L} \Pi[m_l] - \Pi[c_l(\cdot, X_l)]c_l(X_l, X_l)^{-1}(f_l(X_l) - f_{l-1}(X_l) - m_l(X_l))$$

- Computational Cost: $\mathcal{O}(\sum_{l=0}^{L} n_l^3)$ instead of $\mathcal{O}(\sum_{l=0}^{L} n_l)$ for MLMC. This is fine for very expensive integrands!

Multilevel Bayesian Quadrature

- Define $f_{-1} = 0$ for simplicity. Assume $GP(m_l, c_l)$ priors on $f_l - f_{l-1}$ and independence across levels, then, given observations

$$f_l(X_l) - f_{l-1}(X_l) = (f_l(x_{l1}) - f_{l-1}(x_{l1}), \ldots, f_l(x_{ln_l}) - f_{l-1}(x_{ln_l}))^\top$$

for all l, the posterior mean is:

$$\hat{\Pi}_{\text{MLBQ}}[f] := \sum_{l=0}^{L} \Pi[m_l] - \Pi[c_l(\cdot, X_l)] c_l(X_l, X_l)^{-1} (f_l(X_l) - f_{l-1}(X_l) - m_l(X_l))$$

- Computational Cost: $\mathcal{O}(\sum_{l=0}^{L} n_l^3)$ instead of $\mathcal{O}(\sum_{l=0}^{L} n_l)$ for MLMC. This is fine for very expensive integrands!

Convergence Guarantees

- $W^{\tau,2}(\mathcal{X})$ is the Sobolev space of smoothness τ (with norm $\|\cdot\|_\tau$).
- Assume we have a "nice" bounded domain, we choose points "nicely", the kernels c_l have smoothness $\alpha_l > d/2$, the means f_l, f_{l-1} have smoothness $\beta_l > d/2$.
- Then, writing $\tau_l = \min(\alpha_l, \beta_l)$, we have:

$$\underbrace{\left|\Pi[f] - \hat{\Pi}_{\text{MLBQ}}[f]\right|}_{\text{integration error}} \leq \sum_{l=1}^{L} a_l \underbrace{\|f_l - f_{l-1}\|_{\tau_l}}_{\text{very small!}} n_l^{-\frac{\tau_l}{d}} + a_0 \underbrace{\|f_0\|_{\tau_0}}_{\textit{large}} \underbrace{n_0^{-\frac{\tau_0}{d}}}_{\text{very small!}}$$

Teckentrup, A. L. (2020). Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems. SIAM-ASA Journal on Uncertainty Quantification, 8(4), 1310–1337.

Wynne, G., Briol, F.-X., & Girolami, M. (2021). Convergence guarantees for Gaussian process means with misspecified likelihoods and smoothness. Journal of Machine Learning Research, 22(123), 1–40.

Convergence Guarantees

- $W^{\tau,2}(\mathcal{X})$ is the Sobolev space of smoothness τ (with norm $\|\cdot\|_\tau$).
- Assume we have a "nice" bounded domain, we choose points "nicely", the kernels c_l have smoothness $\alpha_l > d/2$, the means f_l, f_{l-1} have smoothness $\beta_l > d/2$.
- Then, writing $\tau_l = \min(\alpha_l, \beta_l)$, we have:

$$\underbrace{\left|\Pi[f] - \hat{\Pi}_{\text{MLBQ}}[f]\right|}_{\text{integration error}} \leq \sum_{l=1}^{L} a_l \underbrace{\|f_l - f_{l-1}\|_{\tau_l}}_{\text{very small!}} n_l^{-\frac{\tau_l}{d}} + a_0 \underbrace{\|f_0\|_{\tau_0}}_{\textit{large}} \underbrace{n_0^{-\frac{\tau_0}{d}}}_{\text{very small!}}$$

Teckentrup, A. L. (2020). Convergence of Gaussian process regression with estimated hyper-parameters and applications in Bayesian inverse problems. SIAM-ASA Journal on Uncertainty Quantification, 8(4), 1310–1337.

Wynne, G., Briol, F.-X., & Girolami, M. (2021). Convergence guarantees for Gaussian process means with misspecified likelihoods and smoothness. Journal of Machine Learning Research, 22(123), 1–40.

Optimising the performance of MLBQ

- Denote by $n^{\text{MLBQ}} = (n_0^{\text{MLBQ}}, \ldots, n_L^{\text{MLBQ}})$ the optimal number of samples per level to obtain a fixed worst-case integration error for function in $W^{\tau,2}(\mathcal{X})$. We show that:

$$n_l^{\text{MLBQ}} \propto \left(\frac{\|f_l - f_{l-1}\|_\tau}{C_l}\right)^{\frac{d}{\tau+d}} \text{ where } \begin{cases} \|\cdot\|_\tau \text{ is the norm of } W^{\tau,2}(\mathcal{X}) \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

- Contrast this with MLMC; looks familiar?

$$n_l^{\text{MLMC}} \propto \left(\frac{V_l}{C_l}\right)^{\frac{1}{2}} \text{ where } \begin{cases} V_l = \text{Var}[f_l - f_{l-1}] = \|f_l - f_{l-1}\|_{L_2(\Pi)} \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

Optimising the performance of MLBQ

- Denote by $n^{\text{MLBQ}} = (n_0^{\text{MLBQ}}, \ldots, n_L^{\text{MLBQ}})$ the optimal number of samples per level to obtain a fixed worst-case integration error for function in $W^{\tau,2}(\mathcal{X})$. We show that:

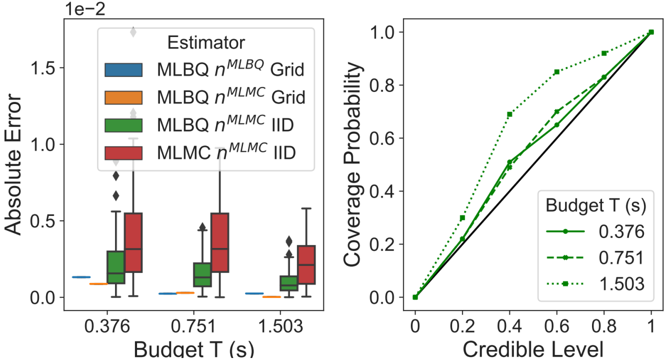
$$n_l^{\text{MLBQ}} \propto \left(\frac{\|f_l - f_{l-1}\|_\tau}{C_l}\right)^{\frac{d}{\tau+d}} \text{ where } \begin{cases} \|\cdot\|_\tau \text{ is the norm of } W^{\tau,2}(\mathcal{X}) \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

- Contrast this with MLMC; looks familiar?

$$n_l^{\text{MLMC}} \propto \left(\frac{V_l}{C_l}\right)^{\frac{1}{2}} \text{ where } \begin{cases} V_l = \text{Var}[f_l - f_{l-1}] = \|f_l - f_{l-1}\|_{L_2(\Pi)} \\ C_l = \text{cost of evaluating } f_l - f_{l-1} \end{cases}$$

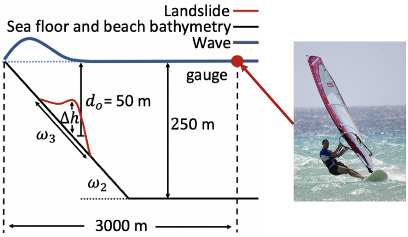
Poisson Equation: A Synthetic Problem

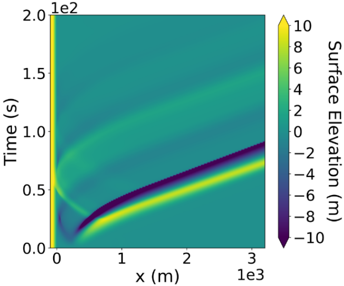
Model: $f''(x) = z(x)$ for $x \in (0,1)$, $f(0) = f(1) = 0$. Integral: $\int_0^1 f(w)dw$.
We use piecewise-linear finite elements with three levels with costs $C = (3.6, 8.5, 42.4)$ (all in 10^{-3} seconds) and Matérn $\frac{1}{2}$ kernel.



Both the number of points per level and placement of the points matters!

Back to Tsunami Models...

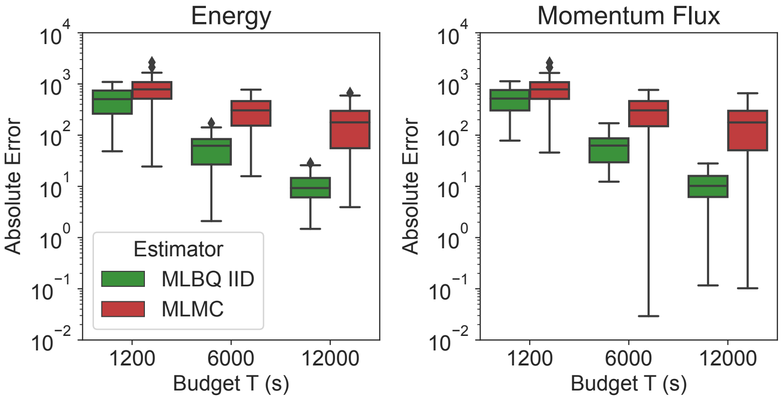




PDE-based tsunami model (non-linear shallow water equations). Need to incorporate uncertainty over parameters.

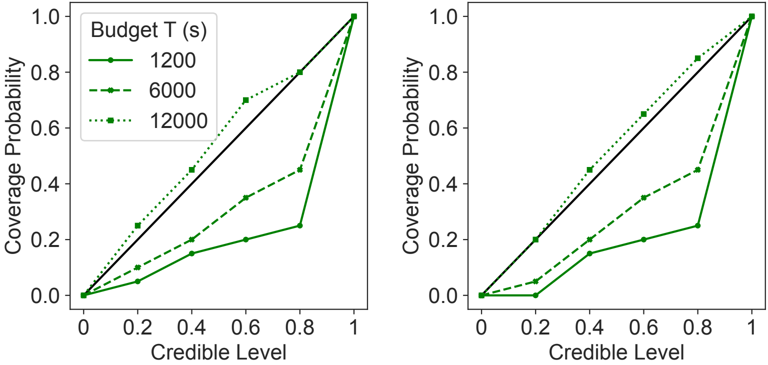
$\approx$ 3 minutes GPU time per integrand evaluations but... very smooth and relatively low-dimensional!

Multilevel Bayesian Quadrature for Tsunami Models



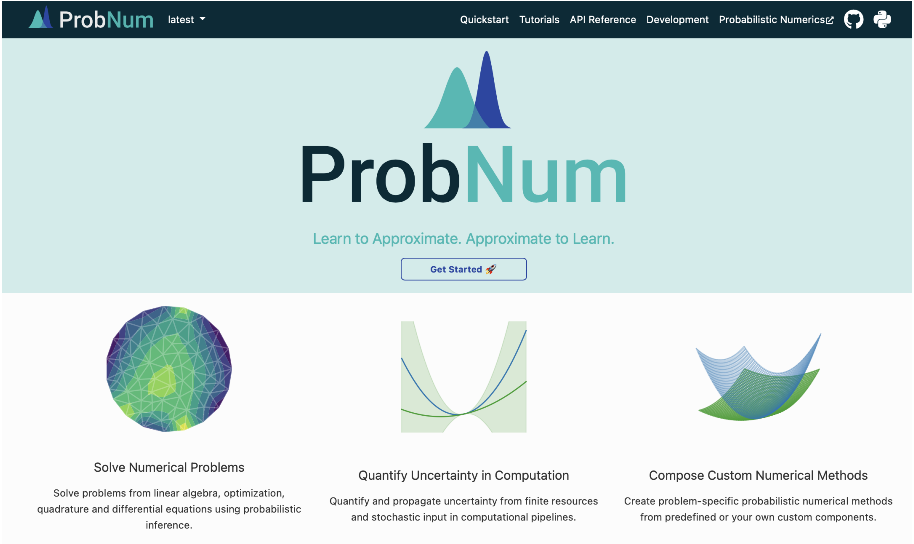
Solver: Volna-OP2, $L = 4$, $C = (5, 15, 30, 65, 160)$ (in seconds), c is a Matérn kernel smoothness $5/2$, $d = 3$.

Multilevel Bayesian Quadrature for Tsunami Models



Conclusion

You Can Try It Out!



Conclusion

- MLMC is great for expensive integration problems, but sometimes you can do even better by using prior information (i.e. MLBQ).
- Bayesian numerical methods can provide an entire posterior distribution over $\Pi[f]$.
  This is useful when there is significant numerical error remaining!
- The method can be combined with many existing approaches (e.g. your favourite point set). There is also clearly scope to extend to adaptive level estimation and refine analysis to specific problems.

Find our more in the paper:

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.

Conclusion

- MLMC is great for expensive integration problems, but sometimes you can do even better by using prior information (i.e. MLBQ).
- Bayesian numerical methods can provide an entire posterior distribution over $\Pi[f]$.
  This is useful when there is significant numerical error remaining!
- The method can be combined with many existing approaches (e.g. your favourite point set). There is also clearly scope to extend to adaptive level estimation and refine analysis to specific problems.

Find our more in the paper:

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.

Conclusion

- MLMC is great for expensive integration problems, but sometimes you can do even better by using prior information (i.e. MLBQ).
- Bayesian numerical methods can provide an entire posterior distribution over $\Pi[f]$.
  This is useful when there is significant numerical error remaining!
- The method can be combined with many existing approaches (e.g. your favourite point set). There is also clearly scope to extend to adaptive level estimation and refine analysis to specific problems.

Find our more in the paper:

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.

Conclusion

- MLMC is great for expensive integration problems, but sometimes you can do even better by using prior information (i.e. MLBQ).
- Bayesian numerical methods can provide an entire posterior distribution over $\Pi[f]$.
  This is useful when there is significant numerical error remaining!
- The method can be combined with many existing approaches (e.g. your favourite point set). There is also clearly scope to extend to adaptive level estimation and refine analysis to specific problems.

Find our more in the paper:

Li, K., Giles, D., Karvonen, T., Guillas, S., & Briol, F.-X. (2023). Multilevel Bayesian quadrature. International Conference on Artificial Intelligence and Statistics (invited for oral presentation), 1845–1868.