[icon]: relevant/new to my research.
[icon]: questions for Lizzie Q...

Multifidelity Survey Paper 7/11

- **Multifidelity methods**: Leverage low-fidelity models for speed up & occasionally use high-fidelity models to course correct / guarantee convergence.
- **General Use of Models**: Describe/characterize input-output relation of system of interest
  ↳ ex. PDE: $f: \mathcal{Z} \to \mathcal{Y}$ for input $z \in \mathcal{Z}$ & output $y \in \mathcal{Y}$ for $\mathcal{Z} \subseteq \mathbb{R}^d$, $\mathcal{Y} \subseteq \mathbb{R}^{d'}$

  ⇒ f evaluation costs $c \in \mathbb{R}_+$ but ↑ accuracy of output approx

<u>high-fidelity model</u> $f_{hi}: \mathcal{Z} \to \mathcal{Y}$ which estimates the output w/ desired accuracy
- <u>low-fidelity model</u> $f_{lo}: \mathcal{Z} \to \mathcal{Y}$ which estimates output w/ lower accuracy

  ↳ Typically, $c_{lo} \in \mathbb{R}_+ < c_{hi} \in \mathbb{R}_+$.

  ↳ we consider $k \in \mathbb{N}$ low-fidelity models, $f_{lo}^{(1)}, \dots, f_{lo}^{(k)}$ that each represent input-output relationship: $f_{lo}^{(i)}: \mathcal{Z} \to \mathcal{Y}$ w/ $c_{lo}^{(i)}$. $i = 1 \dots k$

- <u>outer-loop application</u>: In each iteration, an input $z \in \mathcal{Z}$ is received, $f(z)$ is computed, result is obtained @ end of outer-loop
  ex. uncertainty propogation, SVRG, inverse-problems, sensitivity analysis.
  ↳ outerloop ⊆ many-query except not all many-query apps have desired outer-loop results
  <u>many-query application</u>: applications that evaluate a model many times
- <u>multi-fidelity methods</u>: ~~uses~~ low-fidelity methods to reduce runtime & recourse w/ high fidelity to preserve acc. (accuracy) of outer-loop result ~ result when only using high-fidelity

  $\boxed{\text{Requirements}}$
  ↳ 1) $f_{lo}^{(1)}, \dots, f_{lo}^{(k)}$ <u>useful</u> approx to high-fidelity model f_{hi}
  2) <u>model management strategy</u>: distributes work among models & provides theoretical <u>guarantee</u> of accuracy/conv. of outer-loop result
     1) must balance model evals among models
     ⇒ 2) <u>guarentee</u> same acc as if only high-fidelity model was used

  ⇒ $\boxed{\text{3 MAIN TYPES}}$
  1) adaption of low-fid w/ info from high-fidelity model
  2) fusion of both low & high-fidelity outputs
  3) filtering of high-fid indicated by low-fid filter

Low-fidelity models
- $\boxed{\text{3 MAIN TYPES}}$
  1. simplified models: coarse grid approx, early stopping criteria, "natural" hierarchies ← CG what? Q: what is this? see source 185 Qs
  2. projection-based models: reduced basis models, proper orthogonal decomp, krylov subspace methods
  3. data-fit models: SVM, interpolation/regression, kriging (?) ← is this?

  derivation of several models w/ diff comp cost & fidelities to estimate <u>same</u> output

  ↳ Low fid <u>derived</u> from high-fid
  ↳ 1. by taking advantage of implementation details & domain expertise. Q: Is this similar to Ishigami func?

  ⇒ ex.
  Coarse-grid discretization: Often can derive f_{lo}s by dropping neglecting/ nonlinear terms

  ↳ 2. NOT dep. on high-fid domain knowleage, STILL $\boxed{\text{problem}}$ req. high-fid implementation details
  Mathematically detemine gov. equations of space and project to low dim subspace to get low-fid models

  SVRx? → ↳ 3. NO KNOWLEDGE OF HIGH-FID MODEL. "Black-box". Just need inputs & outputs
  low-fidelity models constructed by fitting <u>coefficients</u> of LC of basis function via <u>interpolation / regression to inputs</u> & high-fidelity outputs.

- **Model Management**: Elizabeth's work:
  Adaptation ex: correction of low-fid model outputs via updates from high-fid
  Fusion ex. control variate method which red var. (σ^2) by exploiting corr btwn low & high fid model

- Kriging = Gaussian Process Regression

simulated annealing
↓
way to sample from diff dist via Markov Chains

✱ pen sucks... $\mathbb{V}_q[I_{hi}\frac{p}{q}] < \mathbb{V}_p[I_{hi}]$ (m's cancel out)

Multifidelity Survey Paper 7/11

yes way correct ☺ Elizabeth's research & my proof of concept

• **Uncertainty Propogation & MC simulation:**

$\mathbb{E}[f_{hi}] = \int_Z f_{hi}(z)p(z)dz$ $Z \sim p$ ← prob density func (PDF) &

For MC estimator w/ m iid $z_1,\dots,z_m \in Z$,

$\mathbb{V}[f_{hi}] = \mathbb{E}[f_{hi}^2] - \mathbb{E}[f_{hi}]^2 \Rightarrow \bar{S}_m^{hi} = \frac{1}{m}\sum_{i=1}^m f_{hi}(z_i)$, ~~$e(\bar S_m^{hi})$~~

but often only useful estimator for large d, $Z \in \mathbb{R}^d$.

MSE $\boxed{e(\bar S_m^{hi}) = \frac{\mathbb{V}[f_{hi}]}{m}}$ RMSE $= \sqrt{e}$ w/ conv. rate $O(m^{-1/2})$ is slow

Control variate method

Let $m_0 \in \mathbb{N}$ = # high-fid eval & for $f_{lo}^{(i)}$, $i \in \{1,\dots,k\}$ let $m_i \in \mathbb{N}$ be low-fid eval s.t. $0 < m_0 \le m_1 \le \dots \le m_k$ & with m_k iid realizations $z_1,\dots,z_{m_k}$, define MC est

$\bar S_{m_0}^{hi} = \frac{1}{m_0}\sum_{j=1}^{m_0} f_{hi}(z_j)$, $\bar S_{m_i}^{(i)} = \frac{1}{m_i}\sum_{j=1}^{m_i} f_{lo}^{(i)}(z_j)$, $\bar S^{MF} = \bar S_{m_0}^{hi} + \sum_{i=1}^k \alpha_i(\bar S_{m_i}^{(i)} - \bar S_{m_{i-1}}^{(i)})$

↳ For $\bar S^{MF}$, $e(\bar S^{MF}) = \mathbb{V}[\bar S^{MF}]$ as $\mathbb{E}[\bar S^{MF}] = \mathbb{E}[f_{hi}]$ (unbiased). & ∴ $c(\bar S^{MF}) = m^T c$

where $m = [m_0,\dots,m_k]^T$ & $c = [c_{hi}, c_{lo}^{(1)},\dots,c_{lo}^{(k)}]^T$

↳ variance reduction:

$$\frac{e(\bar S^{MF})}{e(\bar S^{MC})} = \left(\sqrt{1-\rho_1^2} + \sum_{i=1}^k \sqrt{\frac{c_{lo}^{(i)}}{c_{hi}}(\rho_i^2-\rho_{i+1}^2)}\right)^2$$ where

$\bar S^{MF}$ is computationally cheaper* if $\sqrt{1-\rho_1^2} + \sum_{i=1}^k \sqrt{\frac{c_{lo}^{(i)}}{c_{hi}}(\rho_i^2-\rho_{i+1}^2)} < 1$ * than MC estimator $\bar S_m^{hi}$

Importance Sampling method   Estimate $Z^{-1}(\Upsilon) = \mathbb{E}[I_{hi}]$

Indicator func: $I_{hi}(z) = \begin{cases}1 & f_{hi}(z)<0\\0 & f_{hi}(z)\ge 0\end{cases}$ w/ $\Upsilon = \{z\in Z \mid I_{hi}(z)=1\}$

where $Z\sim p$. Set $Z' \sim q$ where for RV Z_1 Z' is "biasing" RV

S Define support of density p as $supp(p) = \{z \in Z : p(z) > 0\}$

E ∴ For $supp(p) \subset supp(q)$:

E P $\mathbb{E}_p[I_{hi}] = \int_Z I_{hi}(z)p(z)dz \overset{\text{change of var}}{=} \int_{Z'} I_{hi}(z')q(z')\frac{p(z')}{q(z')}dz' = \mathbb{E}_q[I_{hi}\frac{p}{q}]$ (MSE)

A For Imp Samp est, $\bar S_m^{IS} = \frac{1}{m}\sum_{i=1}^m I_{hi}(z_i')p(z_i')/q(z_i')$

P E $\Rightarrow e(\bar S_m^{IS}) = \mathbb{V}_q[I_{hi}\frac{p}{q}]/m$ ✱ better in var red than MC est ∴ if $\mathbb{V}_q[I_{hi}\frac{p}{q}] < \mathbb{V}_p[I_{hi}]$

R For low fidelity, indicator func:

see source 185 $I_{lo}(z) = \begin{cases}1 & f_{lo}(z)<0\\0& f_{lo}(z)\ge0\end{cases}$ & set q s.t. $\{z_i \mid I_{lo}(z_i)=1, i=1,\dots,m\}$

Qs! Then generate Z' from q & eval I_{hi} on z' s.t. we get an estimate of $\mathbb{E}_p[I_{hi}]$. (IF I_{lo} used on Z' then biased est as $\mathbb{E}_p[I_{hi}] \ne \mathbb{E}_p[I_{lo}]$

• **STATISTICAL INFERENCE & ~~Bayesian~~ MCMC framework:** ← MCMC seems closer to my work? (why dis paper so long)

for $f_{hi}: Z \to Y$, let $y_{obs} = f_{hi}(z) + \epsilon$ where ϵ captures "uncertainties" in y_{obs}.

We set $\epsilon \sim N(0,\sigma^2)$ where covariance is denoted as $\Sigma_\epsilon \in \mathbb{R}^{d\times d}$

likelihood function: $\mathcal{L}(y_{obs}|z): Z \to \mathbb{R}$, where $\mathcal{L}(y_{obs}|z) \propto \exp(-\Phi(z))$

data misfit function: $\Phi(z) = \frac12\|\Sigma_\epsilon^{-1/2}(f_{hi}(z)-y_{obs})\|^2$

posterior probability density: $p(z|y_{obs}) \propto \mathcal{L}(y_{obs}|z)p_0(z)$

Metropolis Hastings algorithm Given $\mathcal{L}$, priori density p_0, "proposal" density q, & num of samples m, ← typically $N(z_0)$ where z_0 is sample at prev iteration Q: does this mean q is updated every iter of MH alg?

1) ~~Draw candidate sample, $z^* \sim q(\cdot|z_{i-1})$~~

2) compute acceptance prob, $\alpha(z_{i-1},z^*) = \min\left\{1, \frac{q(z_{i-1}|z^*)\cdot\mathcal{L}(y_{obs}|z^*)\cdot p_0(z^*)}{q(z^*|z_{i-1})\cdot\mathcal{L}(y_{obs}|z_{i-1})\cdot p_0(z_{i-1})}\right\}$

3) Set $z_i = \begin{cases} z^* & \text{w/ prob } \alpha(z_{i-1},z^*)\\ z_{i-1} & \text{w/ prob } 1-\alpha(z_{i-1},z^*)\ \text{(complement)}\end{cases}$

↳ In practice, typically discard first iterations (strongly influenced by priors due to eval of where peak is at) & can't be greedy

For posterior dist w/ funcs of interest $h: Z\to\mathbb{R}$, for $\mathbb{E}[h] = \int_Z h(z)p(z|y_{obs})dz$ can be approx for $z_1,\dots,z_m$ drawn via MCMC alg as $\bar h = \frac1m\sum_{i=1}^m h(z_i)$ Q: $\boxed{\text{how???}}$

↳ can estimate efficiency by effective sample size for a given computational budget wrt $\bar h$

auto correlation time: $\tau_{int}(h) = \frac12 + \sum_{j=1}^\infty \rho_j$ for $\rho_j = corr(h(z_1), h(z_{j+1}))$

$\Rightarrow m_{eff}(h) = m/2\tau_{int}(h)$ s.t. $\mathbb{V}[\bar h] \approx \mathbb{V}[h]/m_{eff}(h)$

MCMC = Markov Chain Monte Carlo
exploitation: min. current f/o
exploration: adapting f/o w/ fns

Multifidelity Survey Paper 7/11

- STATISTICAL INFERENCE & MCMC framework (cont.):
  ↳ Improve efficiency of sampling w/ MCMC:
    1) ↑ m_{eff} for given # MCMC iterations (ex. adaptive MCMC ~~sampling~~)
    2) ↑ # MCMC iter for given comp budget to ↑ m for a given budget (ex. 2-stage MCMC)

  Q: What is the "Hellinger distance" & what does it measure?

- Local & Global Optimization: adaptation model strategy for global & local optimization. must balance exploitation & exploration