Approximation errors, inverse problems and model reduction

Ville Kolehmainen[1]

[1]Department of Applied Physics,
University of Eastern Finland (UEF), Kuopio, Finland

IMA workshop *"Sensor Location in Distributed Parameter System"*, Minneapolis, MN, September 6-8, 2017

Joint work with:

- Antti Lipponen
- Meghdoot Mozumder
- Tanja Tarvainen
- Simon Arridge
- Jari Kaipio

Outline:

Approximation Error Model (AEM)

Approximate marginalization of auxiliary unknowns in inverse problems using the AEM

Construction of low cost predictor model using the AEM

Approximation Error Model (AEM)

Let $y \in \mathbb{R}^m$ and $x \in \mathbb{R}^n$, and let us have:

- An accurate model
$$y = \hat{f}(x)$$
- A reduced model
$$y \approx f(x)$$

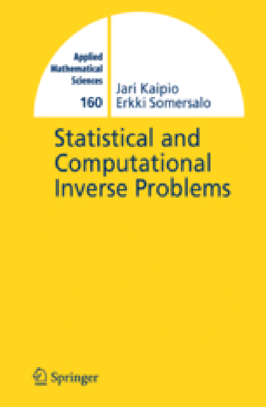
- In the approximation error model (AEM), we write the accurate model as

$$y = f(x) + \underbrace{[\hat{f}(x) - f(x)]}_{\varepsilon(x)} = f(x) + \varepsilon(x)$$

  where $\varepsilon(x)$ is the *approximation error*.

- We discuss how the model can be used for
  1) Marginalization of auxiliary unknowns in inverse problems
  2) Construction of a low cost predictor model for $\hat{f}(x)$.

AEM was originally proposed in:



Approximate marginalization of auxiliary unknowns in inverse problems using the AEM

- Consider the inverse problem of estimating $x \in \mathbb{R}^n$ from noisy observation $y \in \mathbb{R}^m$, given the model

$$y = \hat{f}(x, z) + e$$

  where

  $x \in \mathbb{R}^n$**:** primary unknown
  $z \in \mathbb{R}^d$**:** uninteresting, *auxiliary unknowns* (in this talk, sensor locations & coupling coefficients)

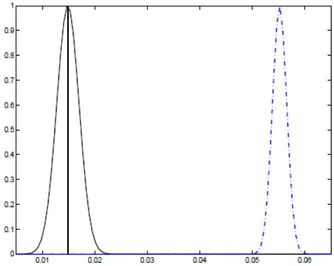
- Complete Bayesian solution: Posterior density model $\pi(x, z|y)$. In many practical applications
  - estimation of all parameters (x, z) or
  - marginalization $\pi(x|y) = \int\int \pi(x, z|y) dz$

  is infeasible due to computation time limitations.

- Classical "ignorance" solution: treat z as fixed conditioning variables, and estimate x from

$$\pi(x|y, z = z_0)$$

  $\rightarrow$ *large errors if z_0 is incorrect*.
- Figure: 1D-marginal posterior $\pi(x_\ell|y)$:
  - exact marginal $\pi(x_\ell|y)$ (black line)
  - $\pi(x_\ell|y, z = z_0)$ with incorrect z_0 (blue)
  - true value of x_ℓ (vertical).



Conventional measurement error model (CEM)

- Consider the conventional measurement model

$$y = \hat{f}(x) + e \tag{1}$$

- Joint density
$\pi(y, x, e) = \pi(y \,|\, x, e)\pi(e \,|\, x)\pi(x) = \pi(y, e \,|\, x)\pi(x)$
- In case of (1), we have $\pi(y \,|\, x, e) = \delta(y - \hat{f}(x) - e)$, and

$$\begin{aligned}
\pi(y \,|\, x) &= \int \pi(y, e \,|\, x)\, d e \\
&= \int \delta(y - \hat{f}(x) - e)\pi(e \,|\, x)\, d e \\
&= \pi_{e \,|\, x}(y - \hat{f}(x) \,|\, x)
\end{aligned}$$

- In the (usual) case of mutually independent x and e, we have $\pi_{e \,|\, x}(e \,|\, x) = \pi_e(e)$ and

$$\pi(y|x) = \pi_e(y - \hat{f}(x))$$

- Furthermore, if $\pi(e) = \mathcal{N}(e_*, \Gamma_e)$ and $\pi(x) = \mathcal{N}(x_*, \Gamma_x)$, we have

$$\pi(x \mid y) \propto \exp\left(-\frac{1}{2}\left(\|L_e(y - \hat{f}(x) - e_*)\|^2 + \|L_x(x - x_*)\|^2\right)\right),$$

  where $L_e^{\mathcal{T}} L_e = \Gamma_e^{-1}$ and $L_x^{\mathcal{T}} L_x = \Gamma_x^{-1}$.
- MAP estimate with the CEM:

$$\min_x \left\{\|L_e(y - \hat{f}(x) - e_*)\|^2 + \|L_x(x - x_*)\|^2\right\}$$

Approximation error model (AEM)

- Accurate measurement model

$$y = \hat{f}(x, z) + e \tag{2}$$

- Instead of using $\hat{f}(x, z)$ and treating (x, z) as unknowns, we fix $z \leftarrow z_0$ and use a possibly drastically reduced forward model

$$x \mapsto f(x, z_0)$$

- We write the accurate measurement model as

$$\begin{aligned} y &= \hat{f}(x,z) + e \\ &= f(x,z_0) + \left[\hat{f}(x,z) - f(x,z_0)\right] + e \\ &= f(x,z_0) + \varepsilon(x,z) + e \end{aligned} \tag{3}$$

  where $\varepsilon(x,z) = \hat{f}(x,z) - f(x,z_0)$ is the *approximation error*.
- The objective is to formulate posterior model

$$\pi(x|y) \propto \pi(y|x)\pi(x)$$

  using measurement model (3).
- We consider e independent of (x,z).

- Using Bayes formula repeatedly, we get

$$\begin{aligned}\pi(y,x,z,e,\varepsilon) &= \pi(y \mid x,z,e,\varepsilon)\pi(x,z,e,\varepsilon)\\ = \delta(y - f(x,z_0) - e - \varepsilon)&\pi(e,\varepsilon \mid x,z)\pi(z \mid x)\pi(x)\\ &= \pi(y,z,e,\varepsilon \mid x)\pi(x)\end{aligned}$$

- Hence

$$\begin{aligned}\pi(y \mid x) &= \iiiint \pi(y,z,e,\varepsilon \mid x)\,de\,d\varepsilon\,dz\\ &= \int \pi_e(y - f(x,z_0) - \varepsilon)\pi_{\varepsilon|x}(\varepsilon \mid x)\,d\varepsilon\end{aligned}$$

(note: convolution integral w.r.t. ε)

- To get a computationally useful and efficient form, π_e and $\pi_{\varepsilon|x}$ are approximated with Gaussian distributions.

- Let the Gaussian approximation of $\pi(\varepsilon, x)$ be

$$\pi(\varepsilon, x) \propto \exp\left\{-\frac{1}{2}\begin{pmatrix} \varepsilon - \varepsilon_* \\ x - x_* \end{pmatrix}^{\mathcal{T}} \begin{pmatrix} \Gamma_\varepsilon & \Gamma_{\varepsilon x} \\ \Gamma_{x\varepsilon} & \Gamma_x \end{pmatrix}^{-1} \begin{pmatrix} \varepsilon - \varepsilon_* \\ x - x_* \end{pmatrix}\right\}$$

- Hence $\pi(e) = \mathcal{N}(e_*, \Gamma_e), \quad \pi(\varepsilon \mid x) = \mathcal{N}(\varepsilon_{*|x}, \Gamma_{\varepsilon|x})$, where

$$\varepsilon_{*|x} = \varepsilon_* + \Gamma_{\varepsilon x}\Gamma_x^{-1}(x - x_*), \quad \Gamma_{\varepsilon|x} = \Gamma_\varepsilon - \Gamma_{\varepsilon x}\Gamma_x^{-1}\Gamma_{x\varepsilon}$$

- Define $\nu \mid x = e + \varepsilon \mid x, \quad \pi(\nu \mid x) = \mathcal{N}(\nu_{*|x}, \Gamma_{\nu|x})$, where

$$\nu_{*|x} = e_* + \varepsilon_* + \Gamma_{\varepsilon x}\Gamma_x^{-1}(x - x_*), \quad \Gamma_{\nu|x} = \Gamma_e + \Gamma_\varepsilon - \Gamma_{\varepsilon x}\Gamma_x^{-1}\Gamma_{x\varepsilon}$$

- Approximate likelihood

$$\pi(y \mid x) = \mathcal{N}(y - f(x, z_0) - \nu_{*|x}, \Gamma_{\nu|x})$$

- Posterior model

$$\pi(x \mid y) \propto \pi(y \mid x)\pi(x) \propto \exp\left(-\frac{1}{2}V(x)\right)$$

where $V(x)$

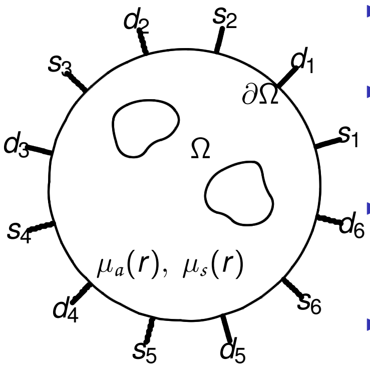
$$\begin{aligned} V(x) &= (y - f(x, z_0) - \nu_{*|x})^{\mathcal{T}} \Gamma_{\nu|x}^{-1} (y - f(x, z_0) - \nu_{*|x}) \\ &\quad + (x - x_*)^{\mathcal{T}} \Gamma_x^{-1} (x - x_*) \\ &= \|L_{\nu|x}(y - f(x, z_0) - \nu_{*|x})\|^2 + \|L_x(x - x_*)\|^2 \end{aligned}$$

where $\Gamma_{\nu\,|\,x}^{-1} = L_{\nu|x}^{\mathcal{T}} L_{\nu|x}$ and $\Gamma_x^{-1} = L_x^{\mathcal{T}} L_x$.

- MAP estimate with the AEM:

$$\min_x \{\|L_{\nu|x}(y - f(x, z_0) - \nu_{*|x})\|^2 + \|L_x(x - x_*)\|^2\}$$

Diffuse Optical Tomography (DOT)



- Imaging of biological tissues using NIR light (turbid medium).
- Target is illuminated at location $s_j \subset \partial\Omega$, response measured at locations $d_k \subset \partial\Omega$.
- Sinusoidally modulated source $\rightarrow$ Amplitude & phase of the modulated wave are measured.
- Inverse problem; Estimate absorption and scattering coefficients $(\mu_a(r), \mu_s(r))$.

Mathematical model

- Diffusion approximation (DA) to the RTE;

$$-\nabla \cdot \kappa(r)\nabla\Phi(r,\omega) + \mu_a(r)\Phi(r,\omega) + \frac{i\omega}{c}\Phi(r,\omega) = 0, \;\; r \in \Omega$$

  where $\kappa(r) = (3(\mu_a(r) + \mu_s(r)))^{-1}$.
- Boundary condition

$$\Phi(r,\omega) + 2\zeta\kappa(r)\frac{\partial\Phi(r,\omega)}{\partial\nu} = g(r,\omega), \;\; r \in \partial\Omega,$$

- Measurable quantity (exitance);

$$\phi = \int_d -\kappa(r)\frac{\partial\Phi(r,\omega)}{\partial\nu}\, dS, \quad d \subset \partial\Omega$$

- Numerical solution by FEM. Notation;

$$y = \hat{f}(x,z), \quad x = (\mu_a, \mu_s)^{\mathcal{T}} \in \mathbb{R}^n$$

Computational Examples

- We consider estimation of $x = (\mu_a, \mu_s)^{\mathcal{T}}$ from

$$y = \hat{f}(x, z) + e$$

- Three cases. Auxiliary parameters z are:
  - **a)** Coupling losses (amplitude & phase shift) of the source and detector fibres.
  - **b)** Locations $\{s_j, d_k\}$ of the sources and detectors.
  - **c)** Combination of a) & b)
- We study the approximate marginalization over a)-c) by the AEM with 2D simulations

Estimates

- **REF**: x estimated using correct realization of z with the conventional error model $y = \hat{f}(x, z) + e$:

$$\min_x \left\{ \|L_e(y - \hat{f}(x, z) - e_*)\|^2 + \|L_x(x - x_*)\|^2 \right\} \qquad (4)$$

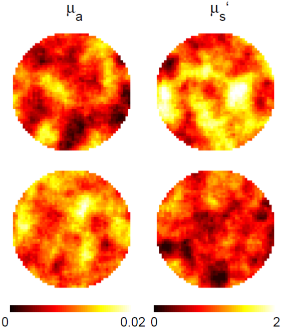
- **CEM**: x estimated using incorrect realization $z = z_0$ and conventional error model $y = f(x, z_0) + e$:

$$\min_x \left\{ \|L_e(y - f(x, z_0) - e_*)\|^2 + \|L_x(x - x_*)\|^2 \right\} \qquad (5)$$

- **AEM**: x estimated using incorrect $z = z_0$ and the approximation error model $y = f(x, z_0) + \varepsilon + e$:

$$\min_x \left\{ \|L_{\nu|x}(y - f(x, z_0) - \nu_{*|x})\|^2 + \|L_x(x - x_*)\|^2 \right\} \qquad (6)$$

Estimation of approximation error statistics



- Approximation error

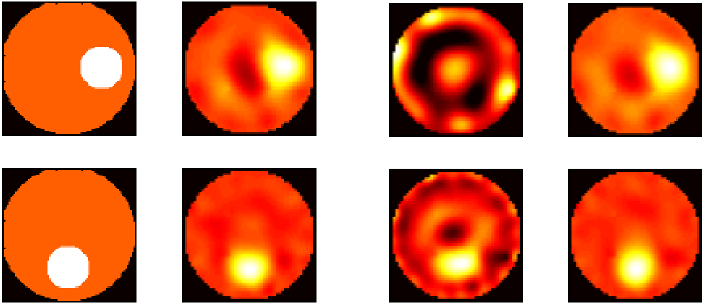
$$\varepsilon(x,z) = \hat{f}(x,z) - f(x,z_0)$$

- Draw sets of samples $\{x^{(\ell)}\}$ and $\{z^{(\ell)}\}$ from $\pi(x)$ and $\pi(z)$
- Compute realizations

$$\varepsilon^{(\ell)} = \hat{f}(x^{(\ell)}, z^{(\ell)}) - f(x^{(\ell)}, z_0)$$

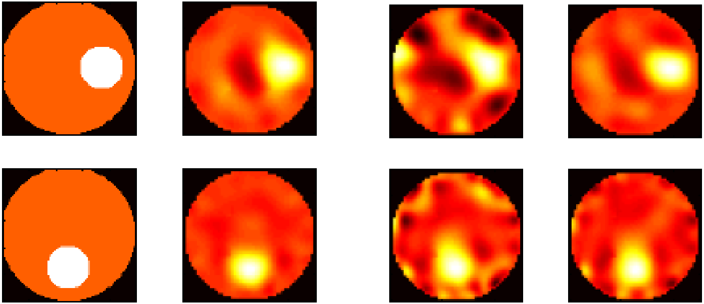
- Estimate ε_* and $\Gamma_{\varepsilon|x}$ as sample averages from $\{x^{(\ell)}, \varepsilon^{(\ell)}\}$.

a) source and detector coupling coefficients



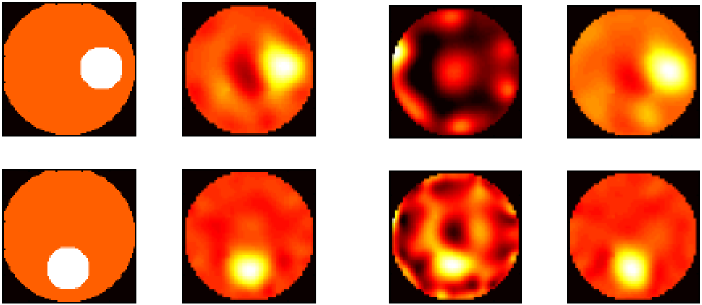
- Top: μ_a. Bottom: μ_s.
- Columns from left to right: 1) True target, 2) REF, 3) CEM, 4) AEM

b) source and detector locations



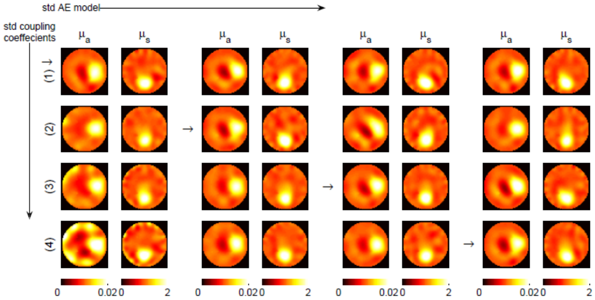
- Top: μ_a. Bottom: μ_s.
- Columns from left to right: 1) True target, 2) REF, 3) CEM, 4) AEM

c) source and detector locations & coupling coefficients



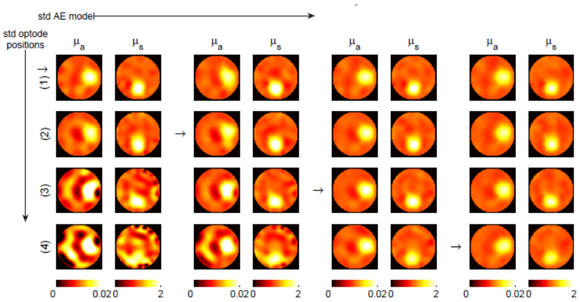
- Top: μ_a. Bottom: μ_s.
- Columns from left to right: 1) True target, 2) REF, 3) CEM, 4) AEM

Robustness w.r.t. the prior model $\pi(z)$ (coupling coefficients)



- Variance of the prior $\pi(z)$ increases from left to right.
- Magnitude of the error $\|z - z_0\|$ increases from top to bottom.

Robustness w.r.t. the prior model $\pi(z)$ (optode locations)



- Variance of the prior $\pi(z)$ increases from left to right.
- Magnitude of the error $\|z - z_0\|$ increases from top to bottom.

Construction of low cost predictor model using the AEM

A low cost predictor model using the AEM:

- Accurate simulation model model

$$y = \hat{f}(x)$$

- Reduced simulation model model

$$y \approx f(x)$$

- (AEM) model

$$y = f(x) + \underbrace{[\hat{f}(x) - f(x)]}_{\varepsilon(x)} = f(x) + \varepsilon(x)$$

A low cost predictor model using the AEM:

- We construct a low cost predictor model

$$\varepsilon(x) \approx P_\varepsilon(x)$$

- Approximate the accurate model by

$$y \approx f(x) + P_\varepsilon(x)$$

- $P_\varepsilon(x)$ constructed using statistical learning.
- Remark: Conventional approach is to construct predictor directly for $\hat{f}(x)$:

$$y \approx P_{\hat{f}}(x)$$

Algorithm:

1. Construct prior model for x
2. Construct training set $\{x^{(\ell)}, \varepsilon(x^{(\ell)})\}$
3. Train/construct the predictor model $P_\varepsilon(x)$ for ε
4. The final simulation model $f(x) + P_\varepsilon(x)$

We evaluate different regressor models in the numerical example.

Simulation models in the numerical example:

- REF: Accurate simulation model $\hat{f}(x)$ used as the reference model.
- RED: Reduced simulation model $f(x)$.
- REG-ACC: Regressor model

  $$P_{\hat{f}}(x)$$

  that predicts the output of the accurate model.
- REG-AE: AE model

  $$f(x) + P_{\varepsilon}(x)$$

  that consists of the reduced model $+$ predictor of the approximation error.

Non-linear heat equation

- Let $x \in [0\ 1]$. We consider solution of

$$\frac{\partial u}{\partial t} - \frac{\partial}{\partial x}\left(\kappa(u)\frac{\partial u}{\partial x}\right) = 0, \tag{7}$$

  where $u = u(x,t)$ is temperature and $\kappa := \kappa(u)$ is the thermal conductivity.
- The initial and boundary conditions

$$u(x,0) = u_0, \tag{8}$$

$$u(0,t) = u(1,t) = 0 \tag{9}$$

  where u_0 is the initial temperature distribution.
- Thermal conductivity

$$\kappa(u) = 0.2 - 0.1/\exp\left(u^2\right).$$

- Let

$$u_{\ell+1} = \hat{f}\,(u_\ell)$$

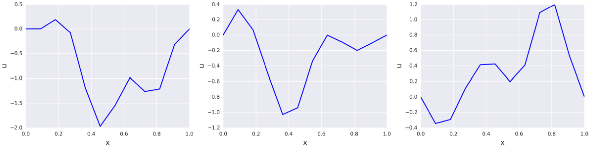
  denote the accurate model (REF) for the solution of the heat equation over an (sampling) interval $\Delta t = 0.01s$ from time index ℓ to $\ell + 1$.
- Accurate model $\hat{f}(u_\ell)$:
  - FD discretization with element size $h = \frac{1}{99}$.
  - Implicit Euler with a time step of $\delta t = \frac{\Delta t}{100}$.
- Reduced model $f(u_\ell)$:
  - FD discretization with element size $h = \frac{1}{11}$.
  - Implicit Euler with a time step of $\delta t = \Delta t$.

Training set for the regressor models

- $\{u_0^{(k)}\}_{k=1}^{N}$ drawn from $u_0 \sim \mathcal{N}(0, \Gamma)$ where

$$\Gamma_{i,j} = \exp\left\{-\frac{|x_i - x_j|^2}{2r^2}\right\} \tag{10}$$

- To emulate temperature distributions at different time instants ℓ, each training sample was scaled as $u^{(k)} = a u_0^{(k)}$, where $a \sim \mathcal{G}amma(0.5, 0.75)$



Three random samples u_0.

Regressor models

- Regressor models:
  - Linear regression
  - Gaussian processes
  - Lasso
  - K nearest neighbors
  - Support Vector Machine
  - Random Forest
- Error metric: the median of the relative error

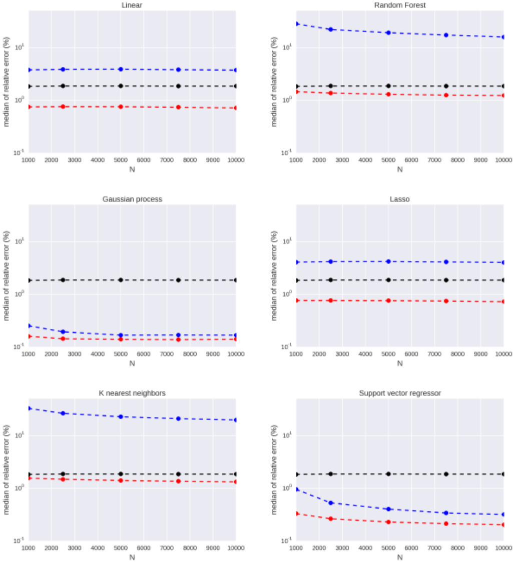
$$\|u_{\ell+1} - f(u_\ell)\| / \|u_{\ell+1}\|$$

Training sample size: Error wrt N using 10-fold cross validation.

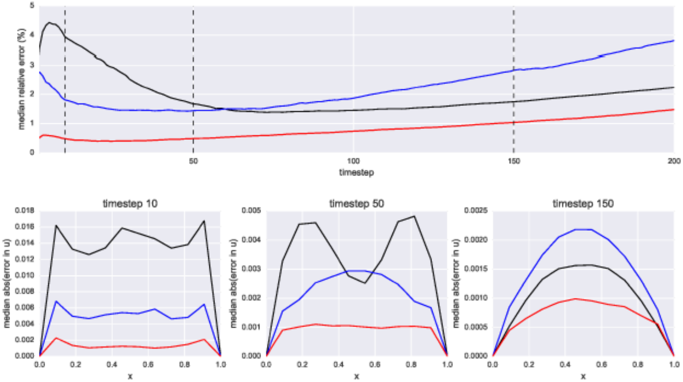
RED: $f(x)$

REG-ACC: $P_{\hat{f}}(x)$

REG-AE: $f(x) + P_{\varepsilon}(x)$



Accuracy of $u(x, t)$ using Gaussian processes ($N = 7500$)



RED: $f(x)$
REG-ACC: $P_{\hat{f}}(x)$
REG-AE: $f(x) + P_{\varepsilon}(x)$

Computation times

Table: Average computation times for different models in heat equation test case corresponding to a simulation of 50 timesteps.

Model	Overall computation time
REF	18.0 s
RED	0.16 s
REG-ACC	0.40 s
REG-AE	0.60 s