Approximate marginalization of uninteresting unknowns in inverse problems

Ville Kolehmainen[1] and Jari Kaipio[2]

[1]Department of Applied Physics,
University of Eastern Finland, Kuopio, Finland
[2]Department of Mathematics, University of Auckland, New Zealand.

SUQ 13, Dunedin, NZ.

Finnish Centre of Excellence
in Inverse Problems Research

UNIVERSITY OF
EASTERN FINLAND

Contents

Part 1: Approximation error model

Introduction

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

$$y = \bar{A}(x, z) + e \in \mathbb{R}^m$$

where

$x \in \mathbb{R}^n$: primary unknown

$z \in \mathbb{R}^d$: uninteresting, auxiliary unknowns (e.g. inaccurately known domain boundary, detector locations, uninteresting distributed parameter, etc)

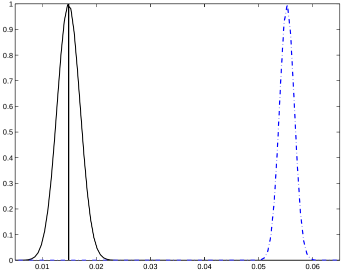
- Complete Bayesian solution: Posterior density model $\pi(x, z|y)$. In many practical applications
  - estimation of all parameters (x, z) or
  - marginalization $\pi(x|y) = \int\int \pi(x, z|y)\mathrm{d}z$

is infeasible due to computation time limitations.

- Conventional (ignorance) solution: treat z as fixed conditioning variables, and estimate x from

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

$\rightarrow$ large errors if realization 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).



- *Approximation error approach* gives approximate (pre)marginalization of $\pi(x, z|y)$ over z;
  - Modeling errors caused by inaccurately known z are modeled as an additive noise process $\varepsilon(x, z)$ in the measurement model.
  - Approximate marginalization over the noise process using a Gaussian approximation for $\pi(x, \varepsilon)$.
  - Approximation of $\pi(x, \varepsilon)$ can be estimated by straightforward Monte Carlo integration over samples from prior models of (x, z). Can be done off-line.
  - Leads to computationally efficient approximation of $\pi(x|y)$, allows simultaneous handling of model reduction related errors.

Conventional measurement error model (CEM)

- Consider the conventional measurement model

$$y = \bar{A}(x) + e \tag{1}$$

- Joint density
$\pi(y, x, e) = \pi(y \mid x, e)\pi(e \mid x)\pi(x) = \pi(y, e \mid x)\pi(x)$
- In case of (1), we have $\pi(y \mid x, e) = \delta(y - \bar{A}(x) - e)$, and

$$\begin{aligned} \pi(y \mid x) &= \int \pi(y, e \mid x)\,\mathrm{d}e \\ &= \int \delta(y - \bar{A}(x) - e)\pi(e \mid x)\,\mathrm{d}e \\ &= \pi_{e \mid x}(y - \bar{A}(x) \mid x) \end{aligned}$$

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

$$\pi(y|x) = \pi_e(y - \bar{A}(x))$$

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

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

  where $L_e^{\mathrm{T}} L_e = \Gamma_e^{-1}$ and $L_x^{\mathrm{T}} L_x = \Gamma_x^{-1}$.

- MAP estimate with the CEM:

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

Approximation error model (AEM)

- Accurate measurement model

$$y = \bar{A}(x, z) + e \in \mathbb{R}^m \qquad (2)$$

- Instead of using (2) and treating (x, z) as unknowns, we fix $z \leftarrow z_0$ and use a possibly drastically reduced model

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

- The use of conventional measurement error model

$$y = A(x, z_0) + e$$

leads to errors in the estimates of x if i) z_0 is incorrect or/and ii) model reduction errors are not negligible.

- In the approximation error approach, we write the measurement model

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

  where $\varepsilon(x,z) = \bar{A}(x,z) - A(x,z_0)$ 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

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

- Hence

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

(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

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

- 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

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

- Posterior model

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

where $V(x)$

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

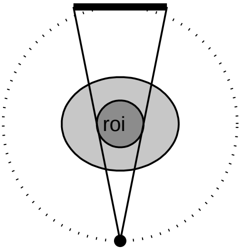
where $\Gamma_{\nu\,|\,x}^{-1} = L_{\nu|x}^{\mathrm{T}} L_{\nu|x}$ and $\Gamma_x^{-1} = L_x^{\mathrm{T}} L_x$.

- MAP estimate with the AEM:

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

Part 2: Computational Examples

Local x-ray tomography



Principle of local tomography.

- Let image domain Ω s.t. $D \subset \Omega$, where D denotes the body. Decomposing x as

$$x = x|_{\Omega \setminus \mathrm{roi}} + x|_{\mathrm{roi}} := x_0 + x_1$$

we get measurement model

$$y = \bar{A}x + e = A_0 x_0 + A_1 x_1 + e \qquad (4)$$

- Estimation of $x = (x_0, x_1)^{\mathrm{T}}$ is often computationally extensive.
- On the other hand, approximating $A_0 x_0 \approx 0$ and estimating only the ROI part x_1 from truncated model

$$y \approx A_1 x_1 + e$$

leads to large estimation errors.

- We write the accurate model as

$$d = A_1 x_1 + \left[\bar{A}x - A_1 x_1\right] + e = A_1 x_1 + \underbrace{\varepsilon + e}_{\nu} \qquad (5)$$

  where

$$\varepsilon = A_0 x_0$$

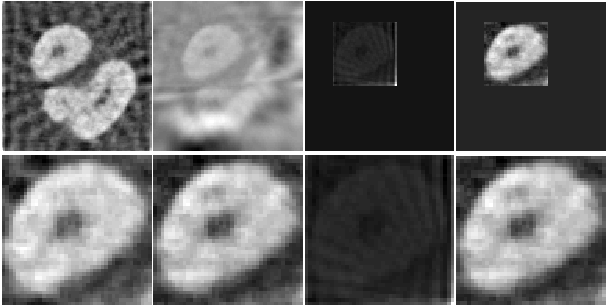
  is the approximation error.
- Gaussian (smoothness) prior $\pi(x) = (x_*, \Gamma_x)$ + linear model $\rightarrow$

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

  has closed form solution.
- MAP estimate

$$\min_{x_1}\{\|L_{\nu|x}(y - A_1 x_1 - \nu_{*|x})\|^2 + \|L_{x_1}(x_1 - x_{*,1})\|^2\}$$

Experimental sparse-angle local tomography data from a tooth specimen



A B C D

Top: Whole image domain Ω, bottom: ROI detail (Ω_1)

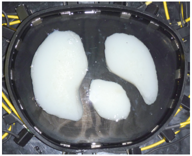
- A Target (reconstruction from global tomography data)
- B Local tomography with accurate projection model and CEM ($y = \bar{A}x + e$)
- C Local tomography with ROI only model and CEM ($y = A_1 x_1 + e$)
- D Local tomography with ROI only model and AEM ($y = A_1 x_1 + \varepsilon + e$)

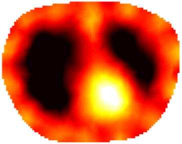
EIT with inaccurately known body shape

- EIT problem: estimate the electrical conductivity $\bar{\sigma}$, given measured voltages $V \in \mathbb{R}^m$ at $\partial\Omega$ and the model

$$V = U(\bar{\sigma}, \gamma) + e, \tag{6}$$

  where $\gamma \in \mathbb{R}^q$ denotes a parameterization of $\partial\Omega$.
- Mapping $U : \bar{\sigma}, \gamma \mapsto U$ is based on FEM approximation of the associated conductivity equation (elliptic PDE).
- Left image: Measurement setup. Right image: Estimated conductivity using correct boundary $\partial\Omega$ in the computational models.

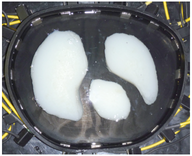
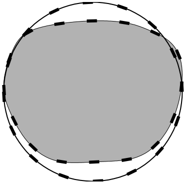
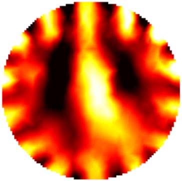




- When EIT is used, for example, in intensive care units, there is no possibility for measuring the exact shape of the patient $\Rightarrow$ Reconstructions are computed using a model domain $\tilde{\Omega}$, i.e., using the model

$$V \approx U(\sigma, \tilde{\gamma}) + e, \tag{7}$$

  where $\tilde{\gamma}$ is parameterization of $\partial\tilde{\Omega}$.
- Middle image: true domain Ω (gray), boundary of $\tilde{\Omega}$ (solid line).
- Right image: estimated conductivity using incorrect boundary $\partial\tilde{\Omega}$.

Approximation error model

- We write measurement model

$$V = U(\sigma,\tilde{\gamma}) + (U(\bar{\sigma},\gamma) - U(\sigma,\tilde{\gamma})) + e = U(\sigma,\tilde{\gamma}) + \varepsilon(\bar{\sigma},\gamma) + e \quad (8)$$

- The relation of conductivities is $\bar{\sigma}(x) = \sigma(T(x))$, where $T(\Omega,\tilde{\Omega}) : \Omega \mapsto \tilde{\Omega}$ is a bijective mapping that models the deformation of domain Ω to $\tilde{\Omega}$.
- T is not unique and not known, and one has to choose a model for the deformation.
- Numerical implementation: $P\bar{\sigma} = \sigma$

Prior models

$\pi(\sigma)$: Proper Gaussian smoothness prior, construction $\sigma(x) = \sigma_{\rm in}(x) + \sigma_{\rm hg}(x)$, where

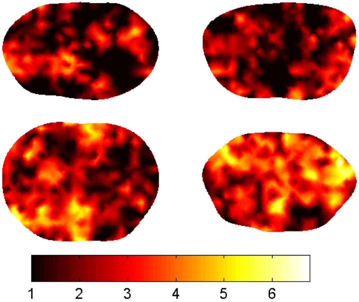
- $\sigma_{\rm in}(x) \sim \mathcal{N}(0, \Gamma_{\rm in})$ (spatially inhomogeneous part)
- $\sigma_{\rm hg}(x) = c\mathbb{I}$, $\mathbb{R} \ni c \sim \mathcal{N}(\sigma_*, \mu_{\rm hg}^2)$ (spatially homogeneous background conductivity).

$\Rightarrow$ $\pi(\sigma) = \mathcal{N}(\sigma_*\mathbb{I}, \Gamma_\sigma)$, where $\Gamma_\sigma = \Gamma_{\rm in} + \mu_{\rm hg}^2 \mathbb{I}\mathbb{I}^{\rm T}$.

$\pi(\gamma)$: Sample based Gaussian $\pi(\gamma) = \mathcal{N}(\gamma_*, \Gamma_\gamma)$, where

- $\gamma_* = \frac{1}{N_{\rm pr}} \sum_{k=1}^{N_{\rm pr}} \gamma^{(k)}$
- $\Gamma_\gamma = \frac{1}{N_{\rm pr}-1} \sum_{k=1}^{N_{\rm pr}} (\gamma^{(k)} - \gamma_*)(\gamma^{(k)} - \gamma_*)^{\rm T}$.
- Sample boundaries $\{\partial\Omega^{(\ell)}, \ell = 1, 2, \ldots, N_{\rm pr}\}$ from chest CT images of $N_{\rm pr} = 150$ different individuals.
- Fourier parameterization of the boundaries.

Estimation of the approximation error statistics



- Sample based approximation for $\pi(\sigma, \varepsilon)$:
  1. Take N_s random draws $\{\gamma^{(\ell)}, \bar{\sigma}^{(\ell)}\}$ from $\pi(\bar{\sigma})$ and $\pi(\gamma)$ (image on the left shows four samples).
  2. Map $\sigma^{(\ell)} = P^{(\ell)}\bar{\sigma}^{(\ell)}$, where $P^{(\ell)} = P^{(\ell)}(\Omega^{(\ell)}, \tilde{\Omega})$ interpolates conductivity from $\Omega^{(\ell)}$ to $\tilde{\Omega}$ according to the deformation model T.
  3. Compute realizations $\varepsilon^{(\ell)} = U(\bar{\sigma}^{(\ell)}, \gamma^{(\ell)}) - U(\sigma^{(\ell)}, \tilde{\gamma})$
  4. Estimate the means and covariances by sample averages.

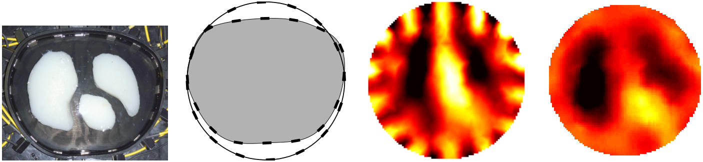
Results

- We make further approximation of modelling σ and ε as mutually independent (i.e, $\Gamma_{\varepsilon\sigma} = 0$) $\Rightarrow$ MAP-AEM estimate

$$\sigma_{\mathrm{MAP}} = \arg\min_{\sigma \geq 0} \left\{ \| L_{e+\varepsilon}(V - U(\sigma,\gamma) - e_* - \varepsilon_*) \|^2 + \| L_\sigma(\sigma - \sigma_*) \|^2 \right\}$$

where $L_{e+\varepsilon}^{\mathrm{T}} L_{e+\varepsilon} = (\Gamma_e + \Gamma_\varepsilon)^{-1}$.

- Images from left to right: a) Measurement setup, b) Ω (gray) and $\partial\tilde{\Omega}$ (solid line), c) MAP-CEM using $\tilde{\Omega}$, d) MAP-AEM using $\tilde{\Omega}$.



Part 3: Modification for high-dimensional data

Modification of AEM for high-dimensional data

- When dimension of data y is large, the computation (and storage) of $\Gamma_{\nu|x}$ and especially $L_{\nu|x}$ can be prohibitive tasks.
- In the following, we make technical approximations:
  - x and ε are approximated mutually independent.
  - Γ_e diagonal
  - We have access to a sample based low rank approximation of Γ_ε (with rank $q < m$).
- Since $\varepsilon - \varepsilon_* \in \mathrm{sp}\{v_1, \ldots, v_q\}$, where v_i are the eigenvectors of Γ_ε, we write

$$\varepsilon = \varepsilon_* + \underbrace{\sum_{k=1}^{p} \alpha_k v_k}_{\varepsilon_p} + \underbrace{\sum_{j=p+1}^{q} \beta_j v_j}_{\varepsilon_{-p}} . \tag{9}$$

- Now, we write

$$\varepsilon_p = \varepsilon_* + \sum_{k=1}^{p} \beta_{p,k} v_k,$$

where $\beta_p \in \mathbb{R}^p$ and $V_p = (v_1, \ldots, v_p) \in \mathbb{R}^{m \times p}$, and write

$$y = A(x) + \varepsilon_p + \varepsilon_{-p} + e = A(x) + V_p\beta_p + \varepsilon_* + \varepsilon_{-p} + e$$

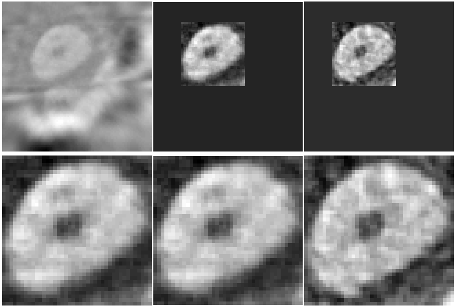
- MAP estimate for the HD modification:

$$\min_{x,\beta_p}\{\|L_{e+\varepsilon_{-p}}(y - A(x) - V_p\beta_p - e_* - \varepsilon_*)\|^2 + \|L_x(x - x_*)\|^2 + \|L_p\beta_p\|^2\}$$

where $L_p = \operatorname{diag}(\lambda_1^{-1/2}, \ldots, \lambda_p^{-1/2})$.

- Low rank approximation for the eigensystem of Γ_ε can be computed with the orthogonal iterations without explicit formation of Γ_ε.
- We can tune dimension p s.t. $L_{e+\varepsilon_{-p}} \approx L_e$.

Local tomography (continued)



A B C

Top: Whole image domain Ω, bottom: ROI detail (Ω_1)

- A Local tomography with accurate projection model and CEM ($y = \bar{A}x + e$)
- B Local tomography with ROI only model and AEM ($y = A_1 x_1 + \varepsilon + e$)
- C Local tomography with ROI only model and HD modification of the AEM ($y = A_1 x_1 + V_p \beta_p + \varepsilon_* + e$)

EIT (continued)

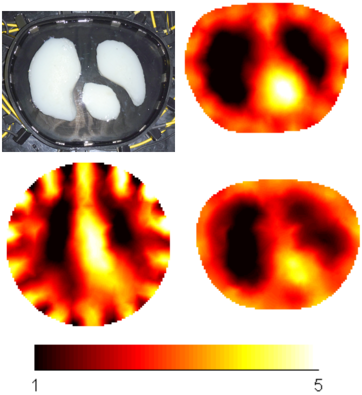
- Using the augmented form, we write

$$V = U(\sigma, \tilde{\gamma}) + V_p \beta_p + \varepsilon_* + \varepsilon_{-p} + e$$

- Once MAP estimate $(\hat{\sigma}, \hat{\beta}_p)$ has been found, we compute estimate for the boundary shape as

$$\hat{\gamma} = \Gamma_{\gamma \varepsilon_p} \Gamma_{\varepsilon_p}^{-1} \hat{\varepsilon}_p + \gamma_*, \quad \hat{\varepsilon}_p = V_p \hat{\beta}_p$$

  where γ_* is the mean of $\pi(\gamma)$.
- Conductivity mapped from $\tilde{\Omega}$ to reconstructed domain by $\hat{\hat{\sigma}} = \tilde{P} \hat{\sigma}$, where $\tilde{P}$ implements interpolation according to the inverse T^{-1} of the domain deformation model.



- Top right: MAP-CEM estimate using the correct domain Ω.
- Bottom left: MAP-CEM using the incorrect model domain $\tilde{\Omega}$.
- Bottom right: MAP-AEM using the incorrect model domain $\tilde{\Omega}$.