Computational Uncertainty Quantification for Inverse Problems: Part 2, Nonlinear Problems

John Bardsley
University of Montana

SIAM Conference on Imaging Science, June 2018

Outline

Computational Uncertainty Quantifications for Inverse Problems

MATLAB codes:
`https://github.com/bardsleyj/SIAMBookCodes`

to be published by SIAM in late-summer/early-fall 2018

- Hierachical Gibbs for Nonlinear Inverse Problems: How do you sample from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$?
- Randomize-then-Optimize as a proposal for Metropolis-Hastings for sampling from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$.
  - some test cases;
  - RTO-MH-within-Hierarchical Gibbs.

Now Consider a Nonlinear Statistical Model

Now assume the non-linear, Gaussian statistical model

$$\mathbf{y} = \mathbf{A}(\mathbf{x}) + \boldsymbol{\epsilon},$$

where

- $\mathbf{y} \in \mathbb{R}^m$ is the vector of observations;
- $\mathbf{x} \in \mathbb{R}^n$ is the vector of unknown parameters;
- $\mathbf{A} : \mathbb{R}^n \to \mathbb{R}^m$ is nonlinear;
- $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I}_m)$, i.e., $\boldsymbol{\epsilon}$ is i.i.d. Gaussian with mean 0 and variance λ^{-1}.

Hierarchical Bayes: Assume Hyper-Priors on λ and δ

Uncertainty in λ and δ: $\lambda \sim p(\lambda)$ and $\delta \sim p(\delta)$. Then

$$p(\mathbf{x}, \lambda, \delta | \mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}, \lambda) p(\lambda)\, p(\mathbf{x}|\delta) p(\delta),$$

is the Bayesian posterior

Hierarchical Bayes: Assume Hyper-Priors on λ and δ

Uncertainty in λ and δ: $\lambda \sim p(\lambda)$ and $\delta \sim p(\delta)$. Then

$$p(\mathbf{x}, \lambda, \delta | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}, \lambda) p(\lambda)\, p(\mathbf{x} | \delta) p(\delta),$$

is the Bayesian posterior, where

$$p(\mathbf{y} | \mathbf{x}, \lambda) \propto \lambda^{m/2} \exp\left(-\frac{\lambda}{2} \|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2\right),$$

and we choose a GMRF prior and Gamma hyper-priors:

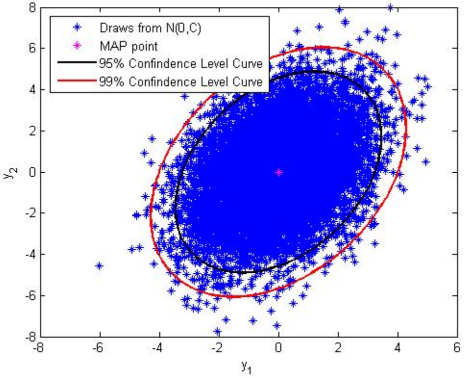
$$\begin{aligned} p(\mathbf{x} | \delta) &\propto \delta^{n/2} \exp\left(-\frac{\delta}{2} \mathbf{x}^T \mathbf{L} \mathbf{x}\right), \\ p(\lambda) &\propto \lambda^{\alpha_\lambda - 1} \exp(-\beta_\lambda \lambda), \\ p(\delta) &\propto \delta^{\alpha_\delta - 1} \exp(-\beta_\delta \delta), \end{aligned}$$

where $\alpha_\lambda = \alpha_\delta = 1$ and $\beta_\lambda = \beta_\delta = 10^{-4}$.

The Full Posterior Distribution: Nonlinear Case

$$p(\mathbf{x}, \lambda, \delta|\mathbf{y}) \propto$$
$$\lambda^{m/2+\alpha_\lambda-1}\delta^{n/2+\alpha_\delta-1}\exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}-\beta_\lambda\lambda-\beta_\delta\delta\right).$$

Sampling versus Computing the MAP Estimator



Hierarchical Gibbs: Nonlinear Case

The full conditionals have the form

$$\begin{aligned} p(\lambda|\mathbf{y},\mathbf{x},\delta) &\propto \lambda^{m/2+\alpha_\lambda-1}\exp\left(-\left(\frac{1}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2+\beta_\lambda\right)\lambda\right); \\ p(\delta|\mathbf{y},\mathbf{x},\lambda) &\propto \delta^{n/2+\alpha_\delta-1}\exp\left(-\left(\frac{1}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}+\beta_\delta\right)\delta\right); \\ p(\mathbf{x}|\mathbf{y},\lambda,\delta) &\propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}\right). \end{aligned}$$

Observations:

1. $p(\mathbf{x}|\mathbf{y},\lambda,\delta)$ is no longer a Gaussian random vector;
2. the product of the first two conditionals is still $p(\lambda,\delta|\mathbf{y},\mathbf{x})$:

$$p(\lambda,\delta|\mathbf{y},\mathbf{x}) = p(\lambda|\mathbf{y},\mathbf{x},\delta)p(\delta|\mathbf{y},\mathbf{x},\lambda);$$

3. and we still have the natural blocking: $(\mathbf{x},\lambda,\delta)=(\mathbf{x};(\lambda,\delta))$.

Sampling from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$: Metropolis-Hastings

Definitions:

$p(\mathbf{x}\|\mathbf{y}, \lambda, \delta)$	posterior (target) density
$\mathbf{x}^k$	Markov chain r.v. at step k
$q(\mathbf{x}\|\mathbf{x}^k)$	proposal density given $\mathbf{x}^k$
$\mathbf{x}^*$	random variable from the proposal

A chain of samples $\{\mathbf{x}^0, \mathbf{x}^1, \cdots\}$ is generated by:

1. Start at $\mathbf{x}^0$
2. For $k = 1, 2, \cdots K$
   - 2.1 sample $\mathbf{x}^* \sim q(\mathbf{x}|\mathbf{x}^{k-1})$
   - 2.2 calculate $r = \min\left\{1, \frac{p(\mathbf{x}^*|\mathbf{y},\lambda,\delta)q(\mathbf{x}^{k-1}|\mathbf{x}^*)}{p(\mathbf{x}^{k-1}|\mathbf{y},\lambda,\delta)q(\mathbf{x}^*|\mathbf{x}^{k-1})}\right\}$
   - 2.3 $\mathbf{x}^k = \begin{cases} \mathbf{x}^* & \text{with probability } r \\ \mathbf{x}^{k-1} & \text{with probability } 1-r \end{cases}$

Metropolis-Hastings Demonstration:

http://chifeng.scripts.mit.edu/stuff/mcmc-demo/

chifeng.scripts.mit.edu/stuff/mcmc-demo/

Randomize-then-Optimize for defining proposal $q_{\text{RTO}}(\mathbf{x})$

First, define

$$\mathbf{A}_{\lambda,\delta}(\mathbf{x}) = \begin{bmatrix} \lambda^{1/2}\mathbf{A}(\mathbf{x}) \\ (\delta\mathbf{L})^{1/2}\mathbf{x} \end{bmatrix} \quad \text{and} \quad \mathbf{y}_{\lambda,\delta} = \begin{bmatrix} \lambda^{1/2}\mathbf{y} \\ \mathbf{0} \end{bmatrix}.$$

Then

$$\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2 + \frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2,$$

and hence

$$p(\mathbf{x}|\mathbf{y},\lambda,\delta) \propto \exp\left(-\frac{1}{2}\|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2\right).$$

Randomized maximum likelihood

Recall: when $\mathbf{A}_{\lambda,\delta}$ is linear, we can sample from $p(\mathbf{x}|\mathbf{y},\lambda,\delta)$ via

$$\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{A}_{\lambda,\delta}(\boldsymbol{\psi}) - (\mathbf{y}_{\lambda,\delta} + \boldsymbol{\epsilon})\|^2, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m+n}).$$

Comment: this is called *randomized maximum likelihood*.

Problem: It is an open question what the probability of $\mathbf{x}$ is.

Extension to nonlinear problems

As in the linear case, we create a nonlinear mapping

$$\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v}), \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n).$$

Extension to nonlinear problems

As in the linear case, we create a nonlinear mapping

$$\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v}), \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n).$$

What are $\mathbf{Q}$ and $\mathbf{F}$? First, define

$$\mathbf{x}_{\lambda,\delta} = \arg\min_{\mathbf{x}} \|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2,$$

then the first-order optimality condition is

$$\mathbf{J}_{\lambda,\delta}(\mathbf{x}_{\lambda,\delta})^T (\mathbf{A}_{\lambda,\delta}(\mathbf{x}_{\lambda,\delta}) - \mathbf{y}_{\lambda,\delta}) = \mathbf{0}.$$

where $\mathbf{J}_{\lambda,\delta}$ is the Jacobian of $\mathbf{A}_{\lambda,\delta}$.

Let $\mathbf{J}_{\lambda,\delta}(\mathbf{x}_{\lambda,\delta}) = \mathbf{QR}$ be the 'thin' $\mathbf{QR}$-factorization, then the first-order optimality condition can be equivalently expressed

$$\mathbf{F}(\mathbf{x}_{\lambda,\delta}) = \mathbf{Q}^T \mathbf{y}_{\lambda,\delta},$$

where $\mathbf{F} \stackrel{\text{def}}{=} \mathbf{Q}^T \mathbf{A}_{\lambda,\delta}$

Let $\mathbf{J}_{\lambda,\delta}(\mathbf{x}_{\lambda,\delta}) = \mathbf{Q}\mathbf{R}$ be the 'thin' $\mathbf{QR}$-factorization, then the first-order optimality condition can be equivalently expressed

$$\mathbf{F}(\mathbf{x}_{\lambda,\delta}) = \mathbf{Q}^T\mathbf{y}_{\lambda,\delta},$$

where $\mathbf{F} \stackrel{\text{def}}{=} \mathbf{Q}^T\mathbf{A}_{\lambda,\delta}$

RTO nonlinear mapping:

$$\begin{aligned} \mathbf{x} &= \mathbf{F}^{-1}\left(\mathbf{Q}^T(\mathbf{y}_{\lambda,\delta} + \boldsymbol{\epsilon})\right), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{m+n}) \\ &\stackrel{\text{def}}{=} \mathbf{F}^{-1}(\mathbf{v}), \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T\mathbf{y}_{\lambda,\delta}, \mathbf{I}_n). \end{aligned}$$

Condition: $\mathbf{F}$ must be invertible with continuous Jacobian.

PDF for $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v})$, $\mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n)$

If $\frac{d}{d\mathbf{x}}\mathbf{F}(\mathbf{x}) = \mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x})$ is invertible for all relevant $\mathbf{x}$,

$$\begin{aligned} q_{\mathrm{RTO}}(\mathbf{x}) &\propto \left| \det\left(\frac{d}{d\mathbf{x}} \mathbf{F}(\mathbf{x}) \right) \right| p_{\mathbf{v}}(\mathbf{F}(\mathbf{x})) \\ &= \left| \det\left(\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x}) \right) \right| \exp\left(-\frac{1}{2} \|\mathbf{Q}^T \mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{Q}^T \mathbf{y}_{\lambda,\delta}\|^2 \right) \end{aligned}$$

PDF for $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v})$, $\mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n)$

If $\frac{d}{d\mathbf{x}}\mathbf{F}(\mathbf{x}) = \mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x})$ is invertible for all relevant $\mathbf{x}$,

$$
\begin{aligned}
q_{\mathrm{RTO}}(\mathbf{x}) &\propto \left|\det\left(\frac{d}{d\mathbf{x}}\mathbf{F}(\mathbf{x})\right)\right| p_{\mathbf{v}}(\mathbf{F}(\mathbf{x})) \\
&= \left|\det\left(\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x})\right)\right| \exp\left(-\frac{1}{2}\|\mathbf{Q}^T \mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{Q}^T \mathbf{y}_{\lambda,\delta}\|^2\right) \\
&= c(\mathbf{x}; \lambda, \delta) p(\mathbf{x}|\mathbf{y}, \lambda, \delta),
\end{aligned}
$$

where

$$
\begin{aligned}
c(\mathbf{x}; \lambda, \delta) &= \left|\det(\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x}))\right| \times \\
&\quad \exp\left(\frac{1}{2}\|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2 - \frac{1}{2}\|\mathbf{Q}^T(\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta})\|^2\right).
\end{aligned}
$$

RTO in Practice: Implementing $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v})$

RTO proposal samples are obtained by solving

$$\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}\|^2, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n),$$

where $\mathbf{F} = \mathbf{Q}^T \mathbf{A}_{\lambda,\delta}$.

RTO in Practice: Implementing $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v})$

RTO proposal samples are obtained by solving

$$\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}\|^2, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n),$$

where $\mathbf{F} = \mathbf{Q}^T \mathbf{A}_{\lambda,\delta}$.

RTO IMPLEMENTATION:

1. Randomize: compute $\mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n)$;
2. Optimize: solve $\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}\|^2$.

RTO in Practice: Implementing $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{v})$

RTO proposal samples are obtained by solving

$$\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}\|^2, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n),$$

where $\mathbf{F} = \mathbf{Q}^T \mathbf{A}_{\lambda,\delta}$.

RTO IMPLEMENTATION:

1. Randomize: compute $\mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n)$;
2. Optimize: solve $\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{F}(\boldsymbol{\psi}) - \mathbf{v}\|^2$.

NOTE: when $\mathbf{A}_{\lambda,\delta}$ is linear, $\mathbf{x} = \mathbf{R}^{-1}\mathbf{v}$, as in the previous slides.

Theorem (RTO probability density)

Let $\mathbf{A}_{\lambda,\delta} : \mathbb{R}^n \to \mathbb{R}^{m+n}$, $\mathbf{y}_{\lambda,\delta} \in \mathbb{R}^{m+n}$, *and assume*

- $\mathbf{A}_{\lambda,\delta}$ *is continuously differentiable;*
- $\mathbf{J}_{\lambda,\delta}(\mathbf{x}) \in \mathbb{R}^{(m+n)\times n}$ *is rank* n *for every* $\mathbf{x}$*;*
- $\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x})$ *is invertible for all relevant* $\mathbf{x}$*.*

Then the random vector defined by

$$\mathbf{x} = \arg\min_{\boldsymbol{\psi}} \|\mathbf{Q}^T \mathbf{A}_{\lambda,\delta}(\boldsymbol{\psi}) - \mathbf{v}\|^2, \quad \mathbf{v} \sim \mathcal{N}(\mathbf{Q}^T \mathbf{y}_{\lambda,\delta}, \mathbf{I}_n),$$

has probability density function

$$q_{\mathrm{RTO}}(\mathbf{x}; \lambda, \delta) \propto c(\mathbf{x}; \lambda, \delta) p(\mathbf{x}|\mathbf{y}, \lambda, \delta),$$

where

$$\begin{aligned} c(\mathbf{x}; \lambda, \delta) &= \left|\det(\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(\mathbf{x}))\right| \times \\ &\quad \exp\left(\frac{1}{2}\|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2 - \frac{1}{2}\|\mathbf{Q}^T(\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta})\|^2\right). \end{aligned}$$

RTO Metropolis-Hastings

Definitions:

$p(\mathbf{x}\|\mathbf{y}, \lambda, \delta)$	posterior (target) density
$\mathbf{x}^k$	Markov chain r.v. at step k
$q_{\mathrm{RTO}}(\mathbf{x}; \lambda, \delta)$	*RTO (independence) proposal density*
$\mathbf{x}^*$	random variable from the proposal

A chain of samples $\{\mathbf{x}^0, \mathbf{x}^1, \cdots\}$ is generated by:

1. Start at $\mathbf{x}^0$
2. For $k = 1, 2, \cdots K$
   - 2.1 sample $\mathbf{x}^* \sim q_{\mathrm{RTO}}(\mathbf{x}; \lambda, \delta)$ from the *RTO proposal density*
   - 2.2 calculate $r = \min\left\{\frac{p(\mathbf{x}^*|\mathbf{y},\lambda,\delta)\, q_{\mathrm{RTO}}(\mathbf{x}^{k-1};\lambda,\delta)}{p(\mathbf{x}^{k-1}|\mathbf{y},\lambda,\delta)\, q_{\mathrm{RTO}}(\mathbf{x}^*;\lambda,\delta)}, 1\right\}$
   - 2.3 $\mathbf{x}^k = \begin{cases} \mathbf{x}^* & \text{with probability } r \\ \mathbf{x}^{k-1} & \text{with probability } 1-r \end{cases}$

RTO-MH Acceptance Ratio

Given RTO sample $\mathbf{x}^* \sim q_{\text{RTO}}(\mathbf{x};\lambda,\delta)$, accept with probability

$$\begin{aligned} r &= \min\left(1, \frac{p(\mathbf{x}^*|\mathbf{y},\lambda,\delta)q_{\text{RTO}}(\mathbf{x}^{k-1};\lambda,\delta)}{p(\mathbf{x}^{k-1}|\mathbf{y},\lambda,\delta)q_{\text{RTO}}(\mathbf{x}^*;\lambda,\delta)}\right) \\ &= \min\left(1, \frac{p(\mathbf{x}^*|\mathbf{y},\lambda,\delta)c(\mathbf{x}^{k-1};\lambda,\delta)p(\mathbf{x}^{k-1}|\mathbf{y},\lambda,\delta)}{p(\mathbf{x}^{k-1}|\mathbf{y},\lambda,\delta)c(\mathbf{x}^*;\lambda,\delta)p(\mathbf{x}^*|\mathbf{y},\lambda,\delta)}\right) \\ &= \min\left(1, \frac{c(\mathbf{x}^{k-1};\lambda,\delta)}{c(\mathbf{x}^*;\lambda,\delta)}\right), \end{aligned}$$

where recall that

$$\begin{aligned} c(\mathbf{x};\lambda,\delta) &= \left|\det(\mathbf{Q}^T\mathbf{J}_{\lambda,\delta}(\mathbf{x}))\right| \times \\ &\quad \exp\left(\frac{1}{2}\|\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta}\|^2 - \frac{1}{2}\|\mathbf{Q}^T(\mathbf{A}_{\lambda,\delta}(\mathbf{x}) - \mathbf{y}_{\lambda,\delta})\|^2\right). \end{aligned}$$

The RTO-MH Algorithm

1. Choose $\mathbf{x}^0 = \mathbf{x}_{\text{MAP}}$ and number of samples N. Set $k = 1$.
2. Compute an RTO sample $\mathbf{x}^* \sim q_{\text{RTO}}(\mathbf{x}; \lambda, \delta)$.
3. Compute the acceptance probability

$$r = \min\left(1, \frac{c(\mathbf{x}^{k-1}; \lambda, \delta)}{c(\mathbf{x}^*; \lambda, \delta)}\right).$$

4. With probability r, set $\mathbf{x}^k = \mathbf{x}^*$, else set $\mathbf{x}^k = \mathbf{x}^{k-1}$.
5. If $k < N$, set $k = k + 1$ and return to Step 2.

Understanding RTO (thanks to Zheng Wang for these slides)

Consider the simple, scalar 'inverse problem':

$$\underbrace{y}_{\text{observation}} = \overbrace{A(x)}^{\text{forward model}} + \underbrace{\epsilon}_{\text{noise}}, \quad x \sim N(0,1), \quad \epsilon \sim N(0,1)$$

$$\begin{aligned}
\underbrace{p(x|y)}_{\text{posterior}} &\propto \exp\left(-\frac{1}{2}\left(A(x)-y\right)^2\right)\exp\left(-\frac{1}{2}x^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left\| \underbrace{\begin{bmatrix} x \\ A(x) \end{bmatrix}}_{\mathbf{A}_{\lambda,\delta}(x)} - \underbrace{\begin{bmatrix} 0 \\ y \end{bmatrix}}_{\mathbf{y}_{\lambda,\delta}} \right\|^2\right) \\
&\propto \exp\left(-\frac{1}{2}\left\|\mathbf{A}_{\lambda,\delta}(x) - \mathbf{y}_{\lambda,\delta}\right\|^2\right)
\end{aligned}$$

Understanding RTO

Least-squares form:

$p(x|y) \propto$

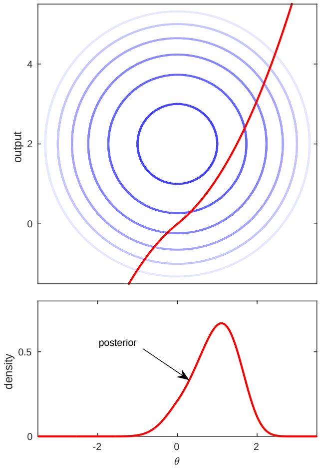
$$\exp\left(-\frac{1}{2}\Big\| \underbrace{\begin{bmatrix} x \\ A(x) \end{bmatrix}}_{\mathbf{A}_{\lambda,\delta}(x)} - \underbrace{\begin{bmatrix} 0 \\ y \end{bmatrix}}_{\mathbf{y}_{\lambda,\delta}} \Big\|^2\right)$$

$p(x|y)$ is the height of the path

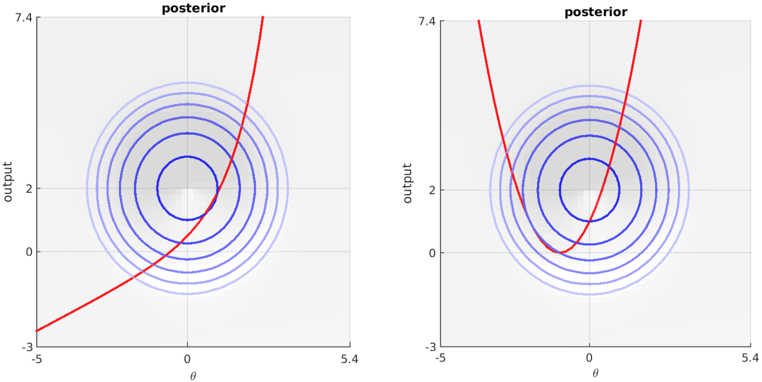
$$\mathbf{A}_{\lambda,\delta}(x) = \begin{bmatrix} x \\ A(x) \end{bmatrix}$$

on the Gaussian

$$\mathcal{N}\left(\mathbf{y}_{\lambda,\delta}, \mathbf{I}_2\right).$$

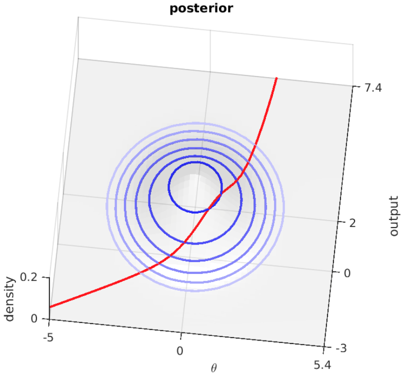


Understanding RTO



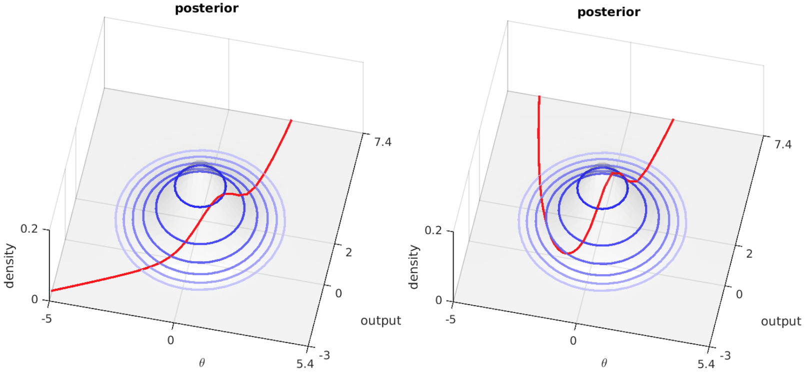
Algorithm's task: sample from the posterior

Understanding RTO



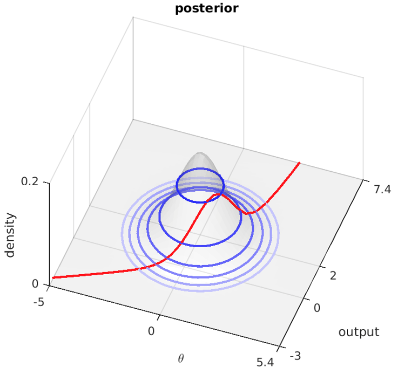
Algorithm's task: sample from the posterior

Understanding RTO



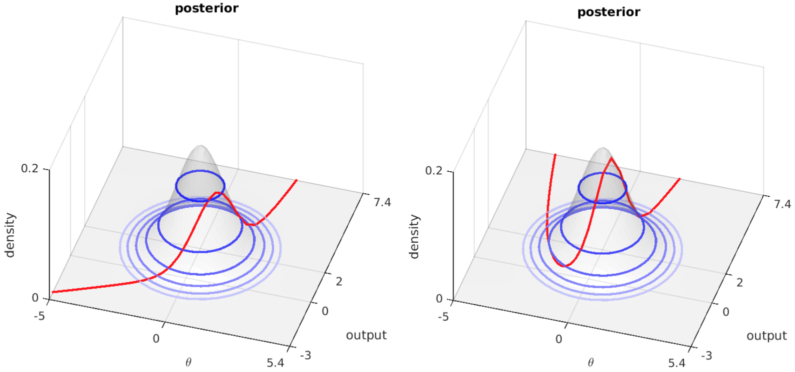
Algorithm's task: sample from the posterior

Understanding RTO



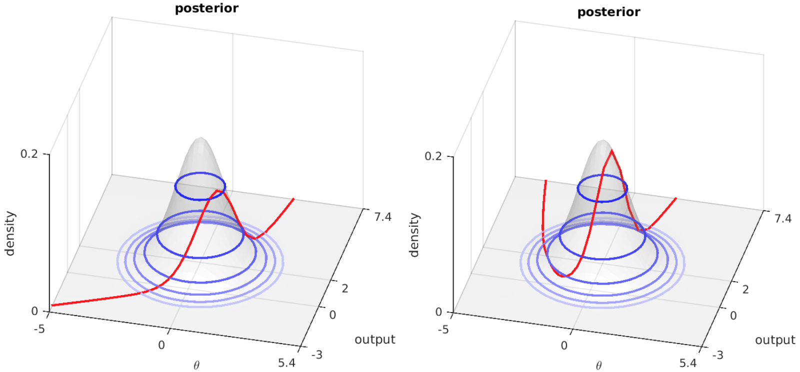
Algorithm's task: sample from the posterior

Understanding RTO



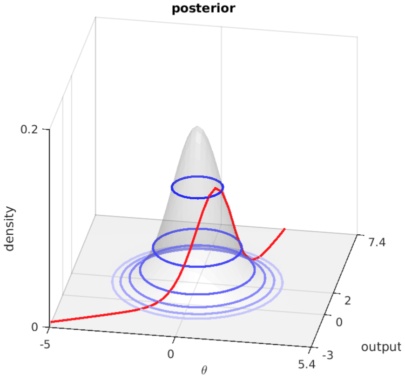
Algorithm's task: sample from the posterior

Understanding RTO



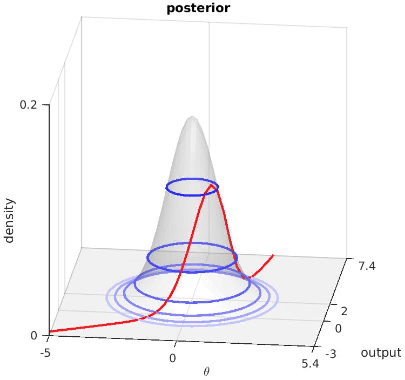
Algorithm's task: sample from the posterior

Understanding RTO



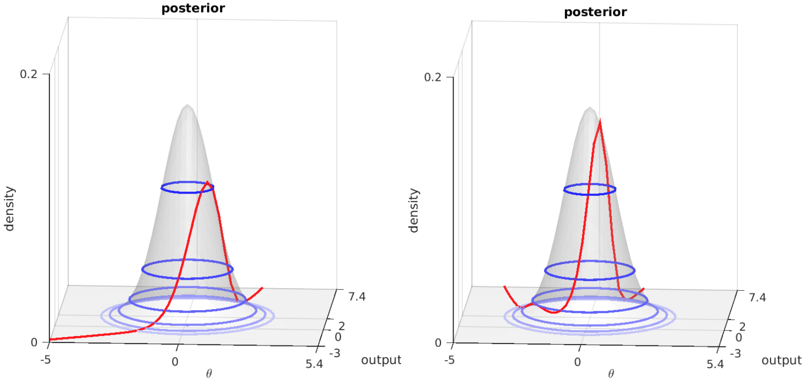
Algorithm's task: sample from the posterior

Understanding RTO



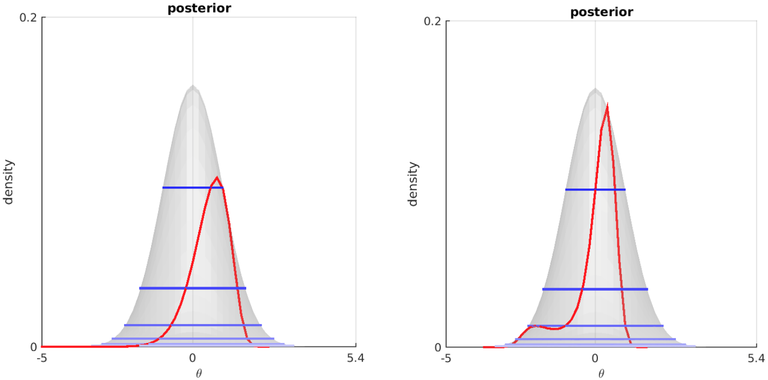
Algorithm's task: sample from the posterior

Understanding RTO



Algorithm's task: sample from the posterior

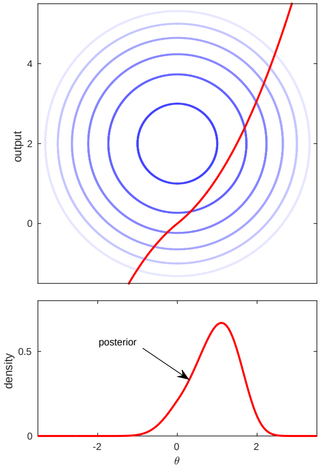
Understanding RTO



Algorithm's task: sample from the posterior

Randomize-then-optimize

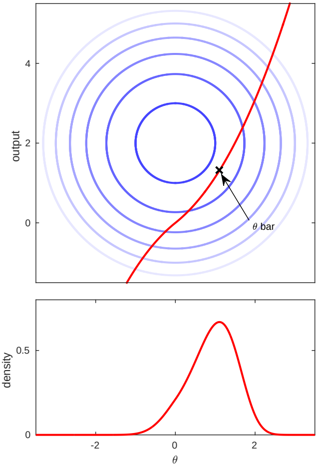
Generate RTO samples $\{x^k\}$:



Randomize-then-optimize

Generate RTO samples $\{x^k\}$:

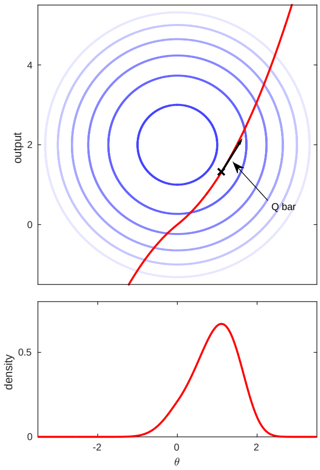
1. Compute $x_{\lambda,\delta}$.



Randomize-then-optimize

Generate RTO samples $\{x^k\}$:

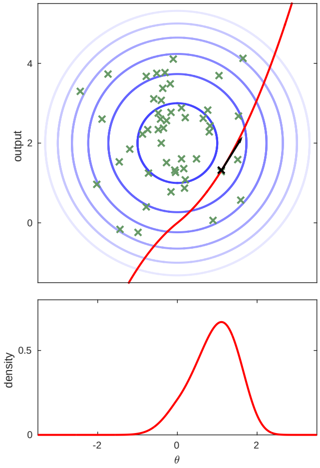
1. Compute $x_{\lambda,\delta}$.
2. Compute $\mathbf{Q} = \mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta}) / \|\mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta})\|$.



Randomize-then-optimize

Generate RTO samples $\{x^k\}$:

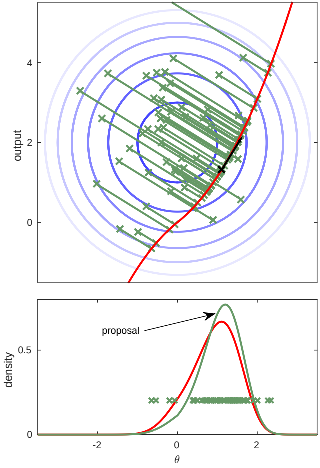
1. Compute $x_{\lambda,\delta}$.
2. Compute $\mathbf{Q} = \mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta})/\|\mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta})\|$.
3. For $k = 1, 2, \cdots, K$
    3.1 Sample $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{y}_{\lambda,\delta}, \mathbf{I}_2)$



Randomize-then-optimize

Generate RTO samples $\{x^k\}$:

1. Compute $x_{\lambda,\delta}$.
2. Compute $\mathbf{Q} = \mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta}) / \|\mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta})\|$.
3. For $k = 1, 2, \cdots, K$
   - 3.1 Sample $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{y}_{\lambda,\delta}, \mathbf{I}_2\right)$
   - 3.2 Compute $x^k = \arg\min_x \left\|\mathbf{Q}^T\left(\mathbf{A}_{\lambda,\delta}(x) - \boldsymbol{\xi}\right)\right\|^2$.



Randomize-then-optimize

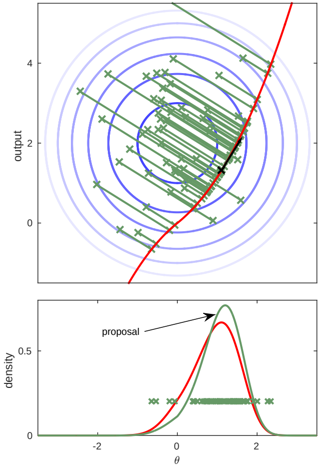
Generate RTO samples $\{x^k\}$:

1. Compute $x_{\lambda,\delta}$.
2. Compute $\mathbf{Q} = \mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta}) / \|\mathbf{J}_{\lambda,\delta}(x_{\lambda,\delta})\|$.
3. For $k = 1, 2, \cdots, K$
   - 3.1 Sample $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{y}_{\lambda,\delta}, \mathbf{I}_2)$
   - 3.2 Compute $x^k = \arg\min_x \left\|\mathbf{Q}^T (\mathbf{A}_{\lambda,\delta}(x) - \boldsymbol{\xi})\right\|^2$.

RTO proposal density:

$$q_{\text{RTO}}(x^k) \propto \left|\mathbf{Q}^T \mathbf{J}_{\lambda,\delta}(x^k)\right|$$

$$\exp\left(-\frac{1}{2}\left\|\mathbf{Q}^T\left(\mathbf{A}_{\lambda,\delta}(x^k) - \mathbf{y}_{\lambda,\delta}\right)\right\|^2\right)$$



Poisson Equation Inverse Problem

Estimate the diffusion coefficient $x(s)$ from measurements of the solution $u(s)$ of

$$-\frac{d}{ds}\left(x(s)\frac{du}{ds}\right) = f(s), \quad 0 < s < 1,$$

where $u(0) = u(1) = 0$, which after discretization takes the form

$$\mathbf{B}(\mathbf{x})\mathbf{u} = \mathbf{f}, \qquad \mathbf{B}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{D}^T \text{diag}(\mathbf{x})\mathbf{D}.$$

Poisson Equation Inverse Problem

Estimate the diffusion coefficient $x(s)$ from measurements of the solution $u(s)$ of

$$-\frac{d}{ds}\left(x(s)\frac{du}{ds}\right) = f(s), \quad 0 < s < 1,$$

where $u(0) = u(1) = 0$, which after discretization takes the form

$$\mathbf{B}(\mathbf{x})\mathbf{u} = \mathbf{f}, \qquad \mathbf{B}(\mathbf{x}) \stackrel{\text{def}}{=} \mathbf{D}^T \text{diag}(\mathbf{x})\mathbf{D}.$$

We generate data using two discrete δ-source vectors $\mathbf{f}_1$ and $\mathbf{f}_2$:

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}_{2N-2\times 1}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} \mathbf{B}(\mathbf{x})^{-1}\mathbf{f}_1 \\ \mathbf{B}(\mathbf{x})^{-1}\mathbf{f}_2 \end{bmatrix}_{2N-2\times 1}}_{\mathbf{A}(\mathbf{x})} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \lambda^{-1}\mathbf{I}_{2N-2}).$$

Nonlinear statistical model: $\mathbf{y} = \mathbf{A}(\mathbf{x}) + \boldsymbol{\epsilon}$.

Assume a GRMF prior: $p(\mathbf{x}|\delta) \propto \exp\left(-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}\right)$.

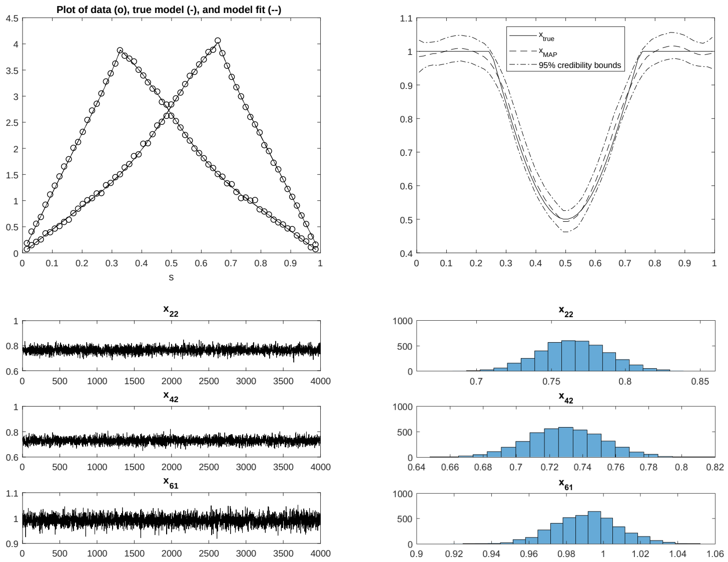
RTO-MH: $p(\mathbf{x}|\mathbf{y},\lambda,\delta) \propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2 - \frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}\right)$

1. Choose $\mathbf{x}^0 = \mathbf{x}_{\mathrm{MAP}}$ and number of samples N. Set $k = 1$.
2. Compute an RTO sample $\mathbf{x}^* \sim q_{\mathrm{RTO}}(\mathbf{x};\lambda,\delta)$.
3. Compute the acceptance probability

$$r = \min\left(1, \frac{c(\mathbf{x}^{k-1};\lambda,\delta)}{c(\mathbf{x}^*;\lambda,\delta)}\right).$$

4. With probability r, set $\mathbf{x}^k = \mathbf{x}^*$, else set $\mathbf{x}^k = \mathbf{x}^{k-1}$.
5. If $k < N$, set $k = k+1$ and return to Step 2.

Numerical Results from RTO-MH Samples

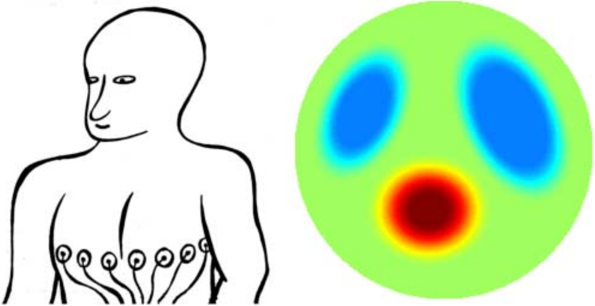


Electrical Impedance Tomography Seppänen, Solonen, Haario, Kaipio

$$\nabla \cdot (x \nabla \varphi) = 0, \qquad \vec{r} \in \Omega$$
$$\varphi + z_\ell x \frac{\partial \varphi}{\partial \vec{n}} = y_\ell, \quad \vec{r} \in e_\ell, \ \ell = 1, \ldots, L$$
$$\int_{e_\ell} x \frac{\partial \varphi}{\partial \vec{n}} \mathrm{d}S = I_\ell, \quad \ell = 1, \ldots, L$$
$$x \frac{\partial \varphi}{\partial \vec{n}} = 0, \qquad \vec{r} \in \partial\Omega \setminus \cup_{\ell=1}^{L} e_\ell$$

- $x = x(\vec{r})$ & $\varphi = \varphi(\vec{r})$: electrical conductivity & potential.
- $\vec{r} \in \Omega$: spatial coordinate.
- e_ℓ: area under the ℓth electrode.
- z_ℓ: contact impedance between ℓth electrode and object.
- y_ℓ & I_ℓ: amplitudes of the electrode potential and current.
- $\vec{n}$: outward unit normal
- L: number of electrodes.

EIT, Forward/Inverse Problem (image by Siltanen)



Left: current $\mathbf{I}$ and electrode potential $\mathbf{y}$; Right: conductivity $\mathbf{x}$.

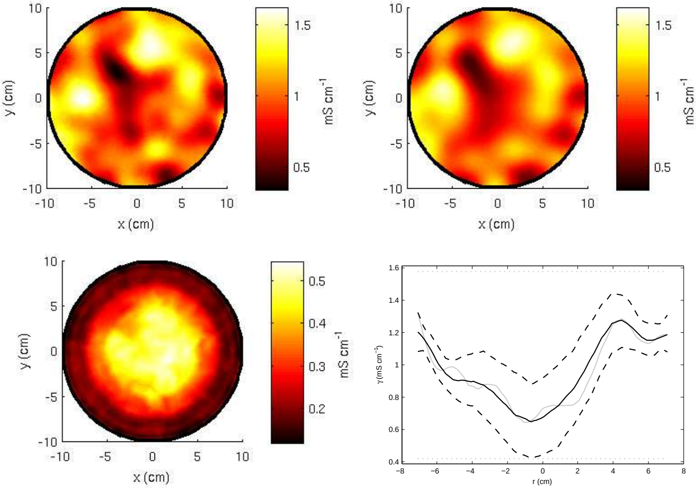
Forward Problem: Given the conductivity $\mathbf{x}$, compute $\mathbf{y} = \mathbf{A}(\mathbf{x})$, where evaluating $\mathbf{A}(\mathbf{x})$ requires solving a complicated PDE.

Inverse Problem: Given $\mathbf{y}$, characterize (sample from)

$$p(\mathbf{x}|\mathbf{y}, \lambda, \delta) \propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x}) - \mathbf{y}\|^2 - \frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}\right).$$

RTO Metropolis-Hastings applied to EIT example

True Conductivity = Realization from Smoothness Prior

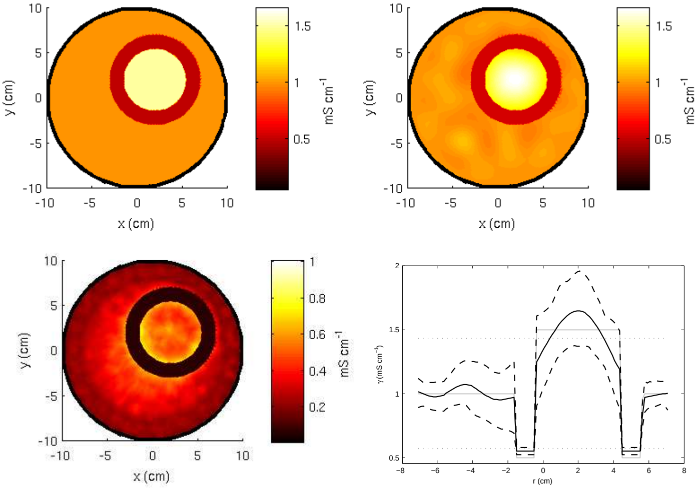


Upper images: truth & conditional mean.

Lower images: 99% c.i.'s & profiles of all of the above.

RTO Metropolis-Hastings applied to EIT example

True Conductivity = Internal Structure #1

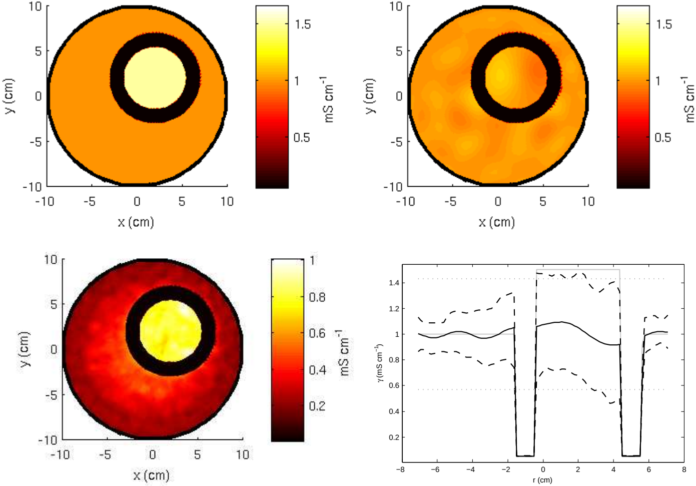


Upper images: truth & conditional mean.

Lower images: 99% c.i.'s & profiles of all of the above.

RTO Metropolis-Hastings applied to EIT example

True Conductivity = Internal Structure #2



Upper images: truth & conditional mean.

Lower images: 99% c.i.'s & profiles of all of the above.

Lognormal Prior and Positivity Constraints

When $\mathbf{x} > 0$, we can assume $\mathbf{x}$ is log-normally distributed:

$$\ln(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, (\delta \mathbf{L})^{-1}),$$

which yields the prior

$$p(\mathbf{x}|\delta) \propto \left(\prod_{i=1}^{n} x_i \right)^{-1} \exp\left(-\frac{\delta}{2} \|\mathbf{L}^{1/2} \ln(\mathbf{x})\|^2 \right).$$

Lognormal Prior and Positivity Constraints

When $\mathbf{x} > 0$, we can assume $\mathbf{x}$ is log-normally distributed:

$$\ln(\mathbf{x}) \sim \mathcal{N}(\mathbf{0}, (\delta \mathbf{L})^{-1}),$$

which yields the prior

$$p(\mathbf{x}|\delta) \propto \left(\prod_{i=1}^{n} x_i \right)^{-1} \exp\left(-\frac{\delta}{2} \|\mathbf{L}^{1/2} \ln(\mathbf{x})\|^2 \right).$$

The transformation $\mathbf{x} = e^{\mathbf{z}}$ then yields

$$\begin{aligned} p(e^{\mathbf{z}}|\mathbf{y}, \lambda, \delta) &\propto p(\mathbf{y}|e^{\mathbf{z}}, \lambda) p(e^{\mathbf{z}}|\delta) \det\left(\frac{d\mathbf{x}}{d\mathbf{z}} \right) \\ &= \exp\left(-\frac{1}{2} \left\| \begin{bmatrix} \lambda^{1/2} \mathbf{A}(e^{\mathbf{z}}) \\ \delta^{1/2} \mathbf{L}^{1/2} \mathbf{z} \end{bmatrix} - \begin{bmatrix} \lambda^{1/2} \mathbf{y} \\ \mathbf{0} \end{bmatrix} \right\|^2 \right). \end{aligned}$$

NOTE: RTO-MH can be used to sample from $p(e^{\mathbf{z}}|\mathbf{y}, \lambda, \delta)$.

RTO-MH to sample from $p(e^{\mathbf{z}}|\mathbf{y}, \lambda, \delta)$

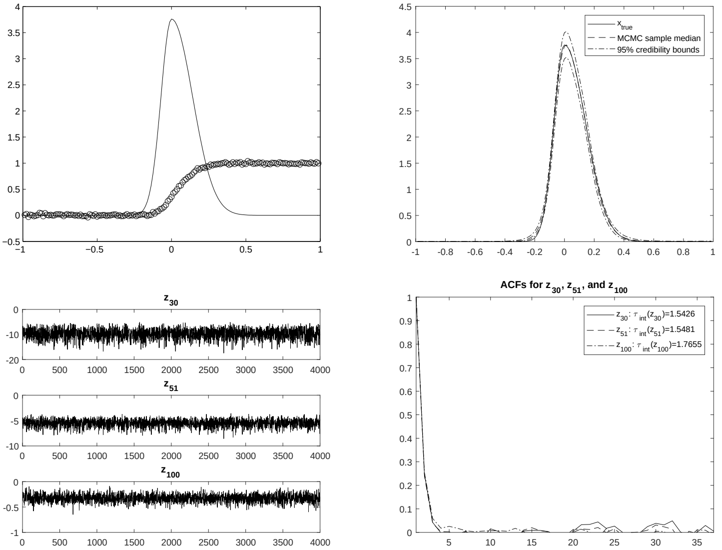
1. Choose $\mathbf{z}^0$, number of samples N, and set $k = 1$.
2. Compute an RTO sample $\mathbf{z}^* \sim q_{\text{RTO}}(\mathbf{z}; \lambda, \delta)$.
3. Compute the acceptance probability

$$r = \min\left(1, \frac{c(\mathbf{z}^{k-1}; \lambda, \delta))}{c(\mathbf{z}^*; \lambda, \delta))}\right).$$

4. With probability r, set $\mathbf{z}^k = \mathbf{z}^*$, else set $\mathbf{z}^k = \mathbf{z}^{k-1}$.
5. If $k < N$, set $k = k + 1$ and return to Step 2.

NOTE: $\left\{\mathbf{x}^k \stackrel{\text{def}}{=} e^{\mathbf{z}^k}\right\}_{k=1}^N$ are samples from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$.

Numerical Test: PSF reconstruction



Laplace/TV/Besov Priors (w/ Zheng, Cui, Marzouk)

Next, we consider the ℓ_1 prior case:

$$p(\mathbf{x}|\delta) \propto \exp\left(-\delta\|\mathbf{D}\mathbf{x}\|_1\right),$$

where $\mathbf{D}$ is an invertible matrix. Then

$$\begin{aligned} p(\mathbf{x}|\mathbf{y},\lambda,\delta) &\propto p(\mathbf{y}|\mathbf{x},\lambda)p(\mathbf{x}|\delta) \\ &= \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2 - \delta\|\mathbf{D}\mathbf{x}\|_1\right). \end{aligned}$$

Note: 1D TV prior & Besov, $B^s_{1,1}$-space, priors have this form.

Laplace/TV/Besov Priors (w/ Zheng, Cui, Marzouk)

Next, we consider the ℓ_1 prior case:

$$p(\mathbf{x}|\delta) \propto \exp\left(-\delta\|\mathbf{D}\mathbf{x}\|_1\right),$$

where $\mathbf{D}$ is an invertible matrix. Then

$$\begin{aligned} p(\mathbf{x}|\mathbf{y},\lambda,\delta) &\propto p(\mathbf{y}|\mathbf{x},\lambda)p(\mathbf{x}|\delta) \\ &= \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2 - \delta\|\mathbf{D}\mathbf{x}\|_1\right). \end{aligned}$$

Note: 1D TV prior & Besov, $B^s_{1,1}$-space, priors have this form.

Problem: $p(\mathbf{x}|\mathbf{y},\lambda,\delta)$ does not have least squares form.

Laplace Prior Transformation

Define the change of variables

$$\mathbf{D}\mathbf{x} = g(\mathbf{z}),$$

such that

$$p(\mathbf{D}^{-1}g(\mathbf{z})|\delta) \propto \exp\left(-\frac{\delta}{2}\|\mathbf{z}\|^2\right)\left|\mathbf{D}^{-1}\mathbf{J}_g(\mathbf{z})\right|.$$

Laplace Prior Transformation

Define the change of variables

$$\mathbf{D}\mathbf{x} = g(\mathbf{z}),$$

such that

$$p(\mathbf{D}^{-1}g(\mathbf{z})|\delta) \propto \exp\left(-\frac{\delta}{2}\|\mathbf{z}\|^2\right)\left|\mathbf{D}^{-1}\mathbf{J}_g(\mathbf{z})\right|.$$

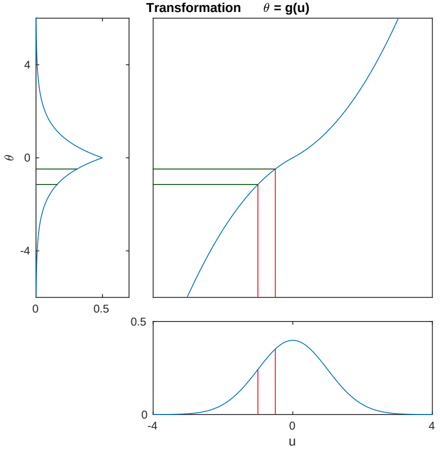
Then the posterior is in least squares form w.r.t. $\mathbf{z}$:

$$\begin{aligned} p(\mathbf{D}^{-1}g(\mathbf{z})|\mathbf{y},\lambda,\delta) &\propto p(\mathbf{y}|\mathbf{D}^{-1}g(\mathbf{z}),\lambda)p(\mathbf{D}^{-1}g(\mathbf{z})|\delta)\det\left(\frac{d\mathbf{x}}{d\mathbf{z}}\right) \\ &= \exp\left(-\frac{\lambda}{2}\|\mathbf{A}\left(\mathbf{D}^{-1}g(\mathbf{z})\right)-\mathbf{y}\|^2-\frac{\delta}{2}\|\mathbf{z}\|^2\right) \\ &= \exp\left(-\frac{1}{2}\left\|\begin{bmatrix}\lambda^{1/2}\left[\mathbf{A}\left(\mathbf{D}^{-1}g(\mathbf{z})\right)-\mathbf{y}\right] \\ \delta^{1/2}\mathbf{z}\end{bmatrix}\right\|^2\right). \end{aligned}$$

Prior Transformation

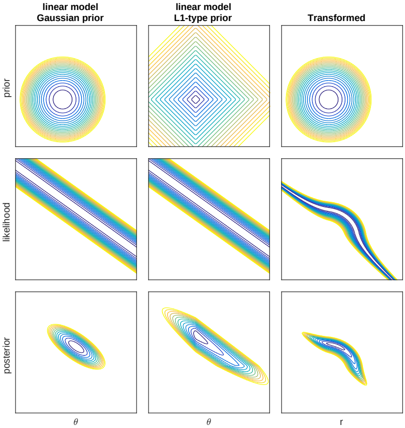
$p(x) \propto$
$\exp(-\lambda|x|)$

$p(z) \propto$
$\exp\left(-\frac{1}{2}z^2\right)$



For multiple independent x_i, transformation is repeated

2D Laplace Prior



Transformation moves complexity from prior to likelihood

RTO-MH to sample from $p(\mathbf{D}^{-1}S(\mathbf{z})|\mathbf{y}, \lambda, \delta)$

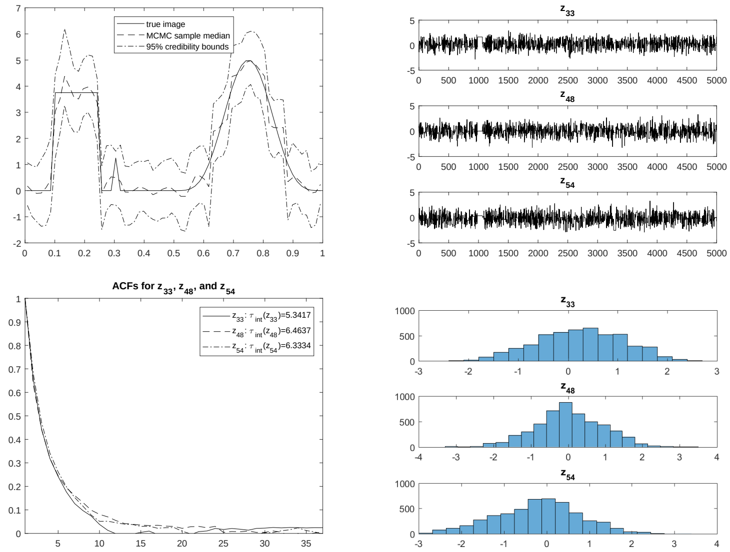
1. Choose $\mathbf{z}^0$, number of samples N, and set $k = 1$.
2. Compute an RTO sample $\mathbf{z}^* \sim q_{\text{RTO}}(\mathbf{z}; \lambda, \delta)$.
3. Compute the acceptance probability

$$r = \min\left(1, \frac{c(\mathbf{z}^{k-1}; \lambda, \delta))}{c(\mathbf{z}^*; \lambda, \delta))}\right).$$

4. With probability r, set $\mathbf{z}^k = \mathbf{z}^*$, else set $\mathbf{z}^k = \mathbf{z}^{k-1}$.
5. If $k < N$, set $k = k + 1$ and return to Step 2.

NOTE: $\left\{\mathbf{x}^k \stackrel{\text{def}}{=} \mathbf{D}^{-1} g(\mathbf{z}^k)\right\}_{k=1}^{N}$ are samples from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$.

Image Deblurring with Besov Space Prior



The Full Posterior Distribution: Nonlinear Case

Recall

$$p(\mathbf{x}, \lambda, \delta|\mathbf{y}) \propto$$

$$\lambda^{m/2+\alpha_\lambda-1}\delta^{n/2+\alpha_\delta-1}\exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}-\beta_\lambda\lambda-\beta_\delta\delta\right).$$

The Full Posterior Distribution: Nonlinear Case

Recall

$$p(\mathbf{x},\lambda,\delta|\mathbf{y}) \propto$$
$$\lambda^{m/2+\alpha_\lambda-1}\delta^{n/2+\alpha_\delta-1}\exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}-\beta_\lambda\lambda-\beta_\delta\delta\right).$$

with full conditionals

$$\begin{aligned}
p(\lambda|\mathbf{y},\mathbf{x},\delta) &\propto \lambda^{m/2+\alpha_\lambda-1}\exp\left(-\left(\frac{1}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2+\beta_\lambda\right)\lambda\right);\\
p(\delta|\mathbf{y},\mathbf{x},\lambda) &\propto \delta^{n/2+\alpha_\delta-1}\exp\left(-\left(\frac{1}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}+\beta_\delta\right)\delta\right);\\
p(\mathbf{x}|\mathbf{y},\lambda,\delta) &\propto \exp\left(-\frac{\lambda}{2}\|\mathbf{A}(\mathbf{x})-\mathbf{y}\|^2-\frac{\delta}{2}\mathbf{x}^T\mathbf{L}\mathbf{x}\right).
\end{aligned}$$

IDEA: use RTO-MH to sample from $p(\mathbf{x}|\mathbf{y},\lambda,\delta)$.

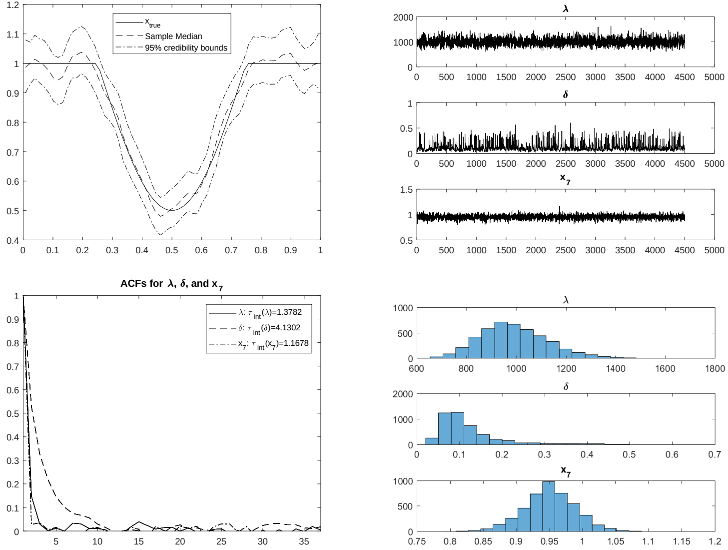
RTO-MH within Hierarchical Gibbs (w/ Cui)

0. Choose $\mathbf{x}^0$, and set $k = 0$;
1. Sample from $p(\lambda, \delta | \mathbf{y}, \mathbf{x}_k)$ via:
   a. $\lambda_{k+1} \sim \Gamma\left(m/2 + \alpha_\lambda, \frac{1}{2}\|\mathbf{A}\mathbf{x}^k - \mathbf{y}\|^2 + \beta_\lambda\right)$;
   b. $\delta_{k+1} \sim \Gamma\left(n/2 + \alpha_\delta, \frac{1}{2}(\mathbf{x}^k)^T \mathbf{L}\mathbf{x}^k + \beta_\delta\right)$;
2. Sample from $p(\mathbf{x}|\mathbf{y}, \lambda_{k+1}, \delta_{k+1})$ using RTO-MH:
   a. compute $\mathbf{x}_* \sim q_{\mathrm{RTO}}(\mathbf{x}; \lambda_{k+1}, \delta_{k+1})$;
   b. set $\mathbf{x}_{k+1} = \mathbf{x}_*$ with probability

$$r = \min\left(1, \frac{c(\mathbf{x}^k; \lambda_{k+1}, \delta_{k+1})}{c(\mathbf{x}^*; \lambda_{k+1}, \delta_{k+1})}\right),$$

   else set $\mathbf{x}_{k+1} = \mathbf{x}_k$.
3. Set $k = k + 1$ and return to step 1.

Test Case: Poisson Equation Inverse Problem



Conclusions/Takeaways

- The development of computationally efficient MCMC methods for nonlinear inverse problems is challenging.
- RTO was presented as a proposal mechanism within Metropolis-Hastings for sampling from $p(\mathbf{x}|\mathbf{y}, \lambda, \delta)$.
- RTO-MH was implemented on several examples:
  - nonlinear inverse problems;
  - lognormal prior (positivity constraints);
  - Laplace/Besov/TV prior;
  - RTO-MH-within-hierarchical Gibbs.