Bayesian Methods and Uncertainty Quantification for Nonlinear Inverse Problems

John Bardsley, University of Montana

Collaborators:
H. Haario, J. Kaipio, M. Laine, Y. Marzouk, A. Seppänen, A. Solonen, Z. Wang

Technical University Denmark, December 2016
Outline

• Nonlinear Inverse Problems Setup
• Randomize-then-Optimize (RTO)
• Test Cases:
  • small # of parameters examples
  • electrical impedance tomography
  • $\ell_1$ priors, i.e., TV and Besov priors
Now consider a nonlinear statistical model

Now assume the non-linear, Gaussian statistical model

\[ y = A(x) + \epsilon, \]

where

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

Consider the following nonlinear, two-parameter **pre-whitened** model.

\[ y_i = x_1(1 - \exp(-x_2t_i)) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2), \quad i = 1, 2, 3, 4, 5, \]

with \( t_i = 2i - 1, \sigma = 0.0136, \) and \( y = [0.076, 0.258, 0.369, 0.492, 0.559]. \)

**GOAL:** estimate a probability distribution for \( x = (x_1, x_2). \)
Toy example continued: the Bayesian posterior $p(x_1, x_2|y)$

$$p(x_1|y) = \int x_2 \, p(x_1, x_2|y) \, dx_2$$

$$p(x_2|y) = \int x_1 \, p(x_1, x_2|y) \, dx_1$$

$p(x_1, x_2|y)$
Markov chain Monte Carlo (MCMC) is a framework for sampling from a (potentially un-normalized) probability distribution.

Some Classical MCMC algorithms

- Gibbs sampling (talk 1: for sampling from $p(x, \lambda, \delta|y)$)
- Metropolis-Hastings
- Adaptive Metropolis (talk 1: for sampling from $p(\lambda, \delta|y)$)

- Inverse Problems: high-dimensional posterior
- Posterior is harder to explore with classical algorithms
- Chains become more correlated, sampling becomes inefficient
Metropolis-Hastings

Definitions:

\[ p(x|y) \] posterior (target) density
\[ x_k \] random variable of the Markov chain at step \( k \)
\[ q(x^*|x_k) \] proposal density given \( x_k \)
\[ x^* \] random variable from the proposal

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

1. Start at \( x^0 \)
2. For \( k = 1, 2, \cdots K \)
   2.1 sample \( x^* \sim q(x^*|x_{k-1}) \)
   2.2 calculate \( \alpha = \min \left\{ \frac{p(x^*|y)q(x_{k-1}|x^*)}{p(x_{k-1}|y)q(x^*|x_{k-1})}, 1 \right\} \)
   2.3 \( x_k = \begin{cases} x^* & \text{with probability } \alpha \\ x_{k-1} & \text{with probability } 1 - \alpha \end{cases} \)
Metropolis-Hastings Demonstration:

http://chifeng.scripts.mit.edu/stuff/mcmc-demo/
Randomize-then-Optimize (RTO): defines a proposal $q$

**Assumption:** RTO requires that the posterior to have least squares form, i.e.,

$$p(x|y) \propto \exp\left(-\frac{1}{2}\|\bar{A}(x) - \bar{y}\|^2\right).$$
Randomize-then-Optimize (RTO): defines a proposal $q$

**Assumption:** RTO requires that the posterior to have least squares form, i.e.,

$$ p(x|y) \propto \exp \left( -\frac{1}{2} \| \bar{A}(x) - \bar{y} \|^2 \right). $$

Given that the likelihood function has the form

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

for which priors $p(x)$ will the posterior density function

$$ p(x|y) \propto p(y|x)p(x). $$

have least squares form?
Test Case 1: Uniform prior

In small parameter cases, it is often true that

\[ p(y|x) = 0 \quad \text{for} \quad x \notin \Omega. \]

Then we can choose as a prior \( p(x) \) defined by

\[ x \sim U(\Omega), \]

where \( U \) denotes the multivariate uniform distribution.
Test Case 1: Uniform prior

In small parameter cases, it is often true that

\[
p(y|x) = 0 \quad \text{for} \quad x \notin \Omega.
\]

Then we can choose as a prior \( p(x) \) defined by

\[
x \sim U(\Omega),
\]

where \( U \) denotes the multivariate uniform distribution.

Then \( p(x) = \text{constant on } \Omega \), and we have

\[
p(x|y) \propto p(y|x)p(x)
\]

\[
\propto \exp\left(-\frac{1}{2}\|A(x) - y\|^2\right).
\]

Thus can use RTO to sample from \( p(x|y) \).
Test Case 2: Gaussian prior

When a Gaussian prior is chosen,

\[ p(x) \propto \exp \left( -\frac{1}{2}(x - x_0)^T L (x - x_0) \right), \]

the posterior can also be written in least squares form:

\[
p(x|y) \propto p(y|x)p(x) \\
\propto \exp \left( -\frac{1}{2} \| A(x) - y \|^2 - \frac{1}{2}(x - x_0)^T L (x - x_0) \right)
\]
Test Case 2: Gaussian prior

When a Gaussian prior is chosen,

\[ p(x) \propto \exp \left( -\frac{1}{2} (x - x_0)^T L (x - x_0) \right) , \]

the posterior can also be written in least squares form:

\[
p(x|y) \propto p(y|x)p(x) \]
\[
\propto \exp \left( -\frac{1}{2} \| A(x) - y \|^2 - \frac{1}{2} (x - x_0)^T L (x - x_0) \right) \]
\[
= \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} A(x) \\ L^{1/2} x \end{bmatrix} - \begin{bmatrix} y \\ L^{1/2} x_0 \end{bmatrix} \right\|^2 \right) \]
\[
def \propto \exp \left( -\frac{1}{2} \| \bar{A}(x) - \bar{y} \|^2 \right) , \]

* Thus we can use RTO to sample from \( p(x|y) \).
Extension of optimization-based approach to nonlinear problems: Randomized maximum likelihood

Recall that when $\bar{A}$ is linear, we can sample from $p(x|y)$ via:

$$x = \arg \min_{\psi} \| \bar{A}(\psi) - (\bar{y} + \epsilon) \|^2, \quad \epsilon \sim \mathcal{N}(0, I_{m+n}).$$

Comment: For nonlinear models, this is called randomized maximum likelihood.

Problem: It is an open question what the probability of $x$ is.
Extension to nonlinear problems

As in the linear case, we create a nonlinear mapping

\[ x = F^{-1}(v), \quad v \sim \mathcal{N}(Q^T\bar{y}, I_n). \]
Extension to nonlinear problems

As in the linear case, we create a nonlinear mapping

\[ x = F^{-1}(v), \quad v \sim \mathcal{N}(Q^T\bar{y}, I_n). \]

What are \( Q \) and \( F \)? First, define

\[ x_{\text{MAP}} = \arg \min_x \| \tilde{A}(x) - \bar{y} \|^2, \]

then first-order optimality yields

\[ J(x_{\text{MAP}})^T (\tilde{A}(x_{\text{MAP}}) - \bar{y}) = 0. \]
So $x_{\text{MAP}}$ is a solution of the nonlinear equation

$$
J(x_{\text{MAP}})^T \bar{A}(x) = J(x_{\text{MAP}})^T \bar{y}.
$$
So $x_{\text{MAP}}$ is a solution of the nonlinear equation

$$J(x_{\text{MAP}})^T \bar{A}(x) = J(x_{\text{MAP}})^T \bar{y}.$$ 

**QR-rewrite:** this equation can be equivalently expressed

$$Q^T \bar{A}(x) = Q^T \bar{y},$$

where $J(x_{\text{MAP}}) = QR$ is the ‘thin’ $QR$ factorization of $J(x_{\text{MAP}})$. 


So $x_{\text{MAP}}$ is a solution of the nonlinear equation

$$J(x_{\text{MAP}})^T \tilde{A}(x) = J(x_{\text{MAP}})^T \tilde{y}. $$

**QR-rewrite:** this equation can be equivalently expressed

$$Q^T \tilde{A}(x) = Q^T \tilde{y},$$

where $J(x_{\text{MAP}}) = QR$ is the ‘thin’ QR factorization of $J(x_{\text{MAP}})$.

**Nonlinear mapping:** define $F \overset{\text{def}}{=} Q^T \tilde{A}$ and

$$x = F^{-1} \left( Q^T (\tilde{y} + \epsilon) \right), \quad \epsilon \sim \mathcal{N}(0, I_{m+n})$$

$$\overset{\text{def}}{=} F^{-1}(v), \quad v \sim \mathcal{N}(Q^T \tilde{y}, I_n).$$
RTO: use optimization to compute \( x = F^{-1}(v) \)

Compute a sample \( x \) from the RTO proposal \( q(x) \):

1. **Randomize**: compute \( v \sim \mathcal{N}(Q^T \bar{y}, I_n) \);
2. **Optimize**: solve

\[
x = \arg \min_{\psi} \left\| F(\psi) - v \right\|^2
\]

3. Reject \( x \) when \( v \) is not in the range of \( F \).
RTO: use optimization to compute $x = F^{-1}(v)$

Compute a sample $x$ from the RTO proposal $q(x)$:
1. **Randomize:** compute $v \sim \mathcal{N}(Q^T\bar{y}, I_n)$;
2. **Optimize:** solve
   $$x = \arg\min_{\psi} \|F(\psi) - v\|^2$$
3. Reject $x$ when $v$ is not in the range of $F$.

**Comment:** steps 1 & 2 can be equivalently expressed

$$x = \arg\min_{\psi} \|Q^T(\tilde{A}(\psi) - (\bar{y} + \epsilon))\|^2, \quad \epsilon \sim \mathcal{N}(0, I_{m+n}).$$
PDF for $x = F^{-1}(v)$, $v \sim \mathcal{N}(Q^T\bar{y}, I_n)$

First, $v \sim \mathcal{N}(Q^T\bar{y}, I_n)$ implies $p_v(v) \propto \exp\left(-\frac{1}{2}\|v - Q^T\bar{y}\|^2\right)$. 
PDF for $x = F^{-1}(v)$, $v \sim \mathcal{N}(Q^T\bar{y}, I_n)$

First, $v \sim \mathcal{N}(Q^T\bar{y}, I_n)$ implies $p_v(v) \propto \exp\left(-\frac{1}{2}\|v - Q^T\bar{y}\|^2\right)$.

Next we need $\frac{d}{dx}F(x) \in \mathbb{R}^{n \times n}$ to be invertible. Then

$$q(x) \propto \left| \det \left( \frac{d}{dx}F(x) \right) \right| p_v(F(x))$$

$$= \left| \det \left( Q^T J(x) \right) \right| \exp \left( -\frac{1}{2}\|Q^T(\bar{A}(x) - \bar{y})\|^2 \right)$$
PDF for $x = F^{-1}(v)$, $v \sim \mathcal{N}(Q^{T}\bar{y}, I_n)$

First, $v \sim \mathcal{N}(Q^{T}\bar{y}, I_n)$ implies $p_{v}(v) \propto \exp \left(-\frac{1}{2}\|v - Q^{T}\bar{y}\|^2\right)$.

Next we need $\frac{d}{dx}F(x) \in \mathbb{R}^{n \times n}$ to be invertible. Then

$$q(x) \propto \left| \det \left( \frac{d}{dx}F(x) \right) \right| p_{v}(F(x))$$

$$= \left| \det \left( Q^{T}J(x) \right) \right| \exp \left(-\frac{1}{2}\|Q^{T}(\bar{A}(x) - \bar{y})\|^2\right)$$

$$= \left| \det \left( Q^{T}J(x) \right) \right| \exp \left(\frac{1}{2}\|\bar{Q}^{T}(\bar{\bar{A}}(x) - \bar{\bar{y}})\|^2\right) \exp \left(-\frac{1}{2}\|\bar{\bar{A}}(x) - \bar{\bar{y}}\|^2\right)$$

$$= c(x)p(x|y),$$

where the columns of $\bar{Q}$ are orthonormal and $C(\bar{Q}) \perp C(Q)$. 
Theorem (RTO probability density)

Let \( \bar{A} : \mathbb{R}^n \to \mathbb{R}^{m+n} \), \( \bar{y} \in \mathbb{R}^{m+n} \), and assume

- \( \bar{A} \) is continuously differentiable;
- \( J(x) \in \mathbb{R}^{(m+n)\times n} \) is rank \( n \) for every \( x \);
- \( Q^T J(x) \) is invertible for all relevant \( x \).

Then the random variable

\[
x = F^{-1} (v), \quad v \sim \mathcal{N}(Q^T \bar{y}, I_n),
\]

has probability density function

\[
q(x) \propto c(x)p(x|y),
\]

where

\[
c(x) = \left| \det(Q^T J(x)) \right| \exp \left( \frac{1}{2} \| Q^T (\bar{y} - \bar{A}(x)) \|^2 \right),
\]

where the columns of \( \bar{Q} \) are orthonormal and \( C(\bar{Q}) \perp C(Q) \).
RTO Metropolis-Hastings

Definitions:

\( p(x|y) \)  posterior (target) density
\( x^k \)  random variable of the Markov chain at step \( k \)
\( q(x^*) \)  RTO (independence) proposal density
\( x^* \)  random variable from the proposal

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

1. Start at \( x^0 \)
2. For \( k = 1, 2, \cdots K \)
   2.1 sample \( x^* \sim q(x^*) \) from the RTO proposal density
   2.2 calculate \( \alpha = \min \left\{ \frac{p(x^*|y)q(x^{k-1})}{p(x^{k-1}|y)q(x^*)}, 1 \right\} \)
   2.3 \( x^k = \begin{cases} x^* & \text{with probability } \alpha \\ x^{k-1} & \text{with probability } 1 - \alpha \end{cases} \)
Metropolis-Hastings using RTO

Given $x^{k-1}$ and proposal $x^* \sim q(x)$, accept with probability

\[ r = \min \left( 1, \frac{p(x^* \mid y)q(x^{k-1})}{p(x^{k-1} \mid y)q(x^*)} \right) \]

\[ = \min \left( 1, \frac{p(x^* \mid y)c(x^{k-1})p(x^{k-1} \mid y)}{p(x^{k-1} \mid y)c(x^*)p(x^* \mid y)} \right) \]

\[ = \min \left( 1, \frac{c(x^{k-1})}{c(x^*)} \right), \]

where recall that

\[ c(x) = \left| \det(Q^T J(x)) \right| \exp \left( \frac{1}{2} \| Q^T (\bar{y} - \bar{A}(x)) \|^2 \right). \]
The RTO Metropolis-Hastings Algorithm

1. Choose $x^0 = x_{\text{MAP}}$ and number of samples $N$. Set $k = 1$.
2. Compute an RTO sample $x^* \sim q(x^*)$.
3. Compute the acceptance probability

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

4. With probability $r$, set $x^k = x^*$, else set $x^k = x^{k-1}$.
5. If $k < N$, set $k = k + 1$ and return to Step 2.
Consider the simple, scalar ‘inverse problem’:

\[
\begin{align*}
y &= f(x) + \epsilon, \quad x \sim N(0, 1), \quad \epsilon \sim N(0, 1) \\
p(x|y) & \propto \exp \left( -\frac{1}{2} (f(x) - y)^2 \right) \exp \left( -\frac{1}{2} x^2 \right) \\
& \propto \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} x \\ f(x) \end{bmatrix} - \begin{bmatrix} 0 \\ y \end{bmatrix} \right\|^2 \right) \\
& \propto \exp \left( -\frac{1}{2} \left\| \bar{\mathbf{A}}(x) - \bar{y} \right\|^2 \right)
\end{align*}
\]
Understanding RTO

Least-squares form:

\[ p(x|y) \propto \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} x \\ f(x) \end{bmatrix} - \begin{bmatrix} 0 \\ y \end{bmatrix} \right\|^2 \right) \]

\[ p(x|y) \] is the height of the path

\[ \bar{A}(x) = [x, f(x)]^T \]

on the Gaussian

\[ \mathcal{N} \left( \begin{bmatrix} 0 \\ y \end{bmatrix}, I_2 \right). \]
Algorithm’s task: sample from the posterior
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
Algorithm’s task: sample from the posterior
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
Algorithm’s task: sample from the posterior
Understanding RTO

Algorithm's task: sample from the posterior
Randomize-then-optimize

Generate RTO samples \( \{x^k\} \):

1. Compute \( x_{\text{MAP}} \).
2. Compute \( Q = J(x_{\text{MAP}}) / \| J(x_{\text{MAP}}) \| \).
3. For \( k = 1, 2, \ldots, K \):
   3.1 Sample \( \xi \sim N(\bar{y}, \mathbb{I}_2) \)
   3.2 Compute \( x^k = \arg\min_x \| Q^T (\bar{A}(x) - \xi) \|_2^2 \).
Randomize-then-optimize

Generate RTO samples \( \{x^k\} \):

1. Compute \( x_{\text{MAP}} \).

\[
\begin{align*}
1. & \text{ Compute } x_{\text{MAP}}. \\
2. & \text{ Compute } Q = J(x_{\text{MAP}}) / \| J(x_{\text{MAP}}) \|. \\
3. & \text{ For } k = 1, 2, \ldots, K \\
   3.1 & \text{ Sample } \xi \sim \mathcal{N}(\bar{y}, I_2). \\
   3.2 & \text{ Compute } x^k = \arg \min_x \| Q^T (\bar{A}(x) - \xi) \|_2.
\end{align*}
\]
Randomize-then-optimize

Generate RTO samples \( \{x^k\} \):

1. Compute \( x_{\text{MAP}} \).

2. Compute \( Q = \frac{J(x_{\text{MAP}})}{\|J(x_{\text{MAP}})\|} \).
Randomize-then-optimize

Generate RTO samples \(\{x^k\}\):

1. Compute \(x_{\text{MAP}}\).
2. Compute \(Q = \frac{J(x_{\text{MAP}})}{\|J(x_{\text{MAP}})\|}\).
3. For \(k = 1, 2, \cdots, K\)
   3.1 Sample \(\xi \sim \mathcal{N}(\bar{y}, I_2)\)
Generate RTO samples \( \{x^k\} \):

1. Compute \( x_{\text{MAP}} \).
2. Compute \( Q = J(x_{\text{MAP}})/\|J(x_{\text{MAP}})\| \).
3. For \( k = 1, 2, \ldots, K \)
   3.1 Sample \( \xi \sim \mathcal{N}(\bar{y}, I_2) \)
   3.2 Compute \( x^k = \arg \min_x \|Q^T(\bar{A}(x) - \xi)\|^2 \).
Randomize-then-optimize

Generate RTO samples \( \{x^k\} \):

1. Compute \( x_{\text{MAP}} \).
2. Compute \( Q = J(x_{\text{MAP}})/\|J(x_{\text{MAP}})\| \).
3. For \( k = 1, 2, \cdots, K \)
   3.1 Sample \( \xi \sim \mathcal{N}(\bar{y}, I_2) \)
   3.2 Compute \( x^k = \arg\min_x \|Q^T(\bar{A}(x) - \xi)\|^2 \).

RTO proposal density:

\[
q(x^k) \propto \left| Q^T J(x^k) \right| \exp \left( -\frac{1}{2} \|Q^T(\bar{A}(x^k) - \bar{y})\|^2 \right)
\]
Choose prior $p(x)$ defined by

$$x \sim U(\Omega),$$

where $U$ is a multivariate uniform distribution on $\Omega$. Then $p(x) = \text{constant on } \Omega$, and we have

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

Thus can use RTO to sample from $p(x|y)$. 
BOD, Good: \( A(x_1, x_2) = x_1(1 - \exp(-x_2 t)) \)

- \( t = 20 \) linearly spaced observations in \( 1 \leq x \leq 9 \);
- \( y = A(x_1, x_2) + \epsilon \), where \( \epsilon \sim \mathcal{N}(0, \sigma^2 I) \) with \( \sigma = 0.01 \);
- \([x_1, x_2] = [1, 0.1]^T\).
BOD, Bad: \( A(x_1, x_2) = x_1(1 - \exp(-x_2 t)) \)

- \( t = 20 \) linearly spaced observations in \( 1 \leq x \leq 5 \);
- \( y = A(x_1, x_2) + \epsilon \), where \( \epsilon \sim \mathcal{N}(0, \sigma^2 I) \) with \( \sigma = 0.01 \);
- \([x_1, x_2] = [1, 0.1]^T\).
MONOD: \( A(x_1, x_2) = \frac{x_1 t}{x_2 + t} \)

\[
\begin{align*}
y &= [0.053, 0.060, 0.112, 0.105, 0.099, 0.122, 0.125]^T .
\end{align*}
\]
Autocorrelation plots for $x_1$ for Good and Bad BOD
Gaussian prior test case

When a Gaussian prior is chosen,

\[ p(x) \propto \exp \left( -\frac{1}{2} (x - x_0)^T L (x - x_0) \right) , \]

the posterior can be written in least squares form:

\[
 p(x|y) \propto \exp \left( -\frac{1}{2} \| A(x) - y \|^2 - \frac{1}{2} (x - x_0)^T L (x - x_0) \right)
\]

\[
 = \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} A(x) \\ L^{1/2} x \end{bmatrix} - \begin{bmatrix} y \\ L^{1/2} x_0 \end{bmatrix} \right\|^2 \right)
\]

\[
 \overset{\text{def}}{=} \exp \left( -\frac{1}{2} \| \bar{A}(x) - \bar{y} \|^2 \right) .
\]

\* Thus we can use RTO to sample from \( p(x|y) \).
\[ \nabla \cdot (x \nabla \varphi) = 0, \quad \vec{r} \in \Omega \]
\[ \varphi + z_\ell x \frac{\partial \varphi}{\partial \vec{n}} = y_\ell, \quad \vec{r} \in e_\ell, \; \ell = 1, \ldots, L \]
\[ \int_{e_\ell} x \frac{\partial \varphi}{\partial \vec{n}} \, dS = I_\ell, \quad \ell = 1, \ldots, L \]
\[ x \frac{\partial \varphi}{\partial \vec{n}} = 0, \quad \vec{r} \in \partial \Omega \setminus \bigcup_{\ell=1}^L e_\ell \]

- \( x = x(\vec{r}) \) & \( \varphi = \varphi(\vec{r}) \): electrical conductivity & potential.
- \( \vec{r} \in \Omega \): spatial coordinate.
- \( e_\ell \): area under the \( \ell \)th electrode.
- \( z_\ell \): contact impedance between \( \ell \)th electrode and object.
- \( y_\ell \) & \( I_\ell \): amplitudes of the electrode potential and current.
- \( \vec{n} \): outward unit normal
- \( L \): number of electrodes.
Forward Problem: Given the conductivity $x$, compute

$$y = f(x) + \epsilon.$$ 

Evaluating $f(x)$ requires solving a complicated PDE.

Inverse Problem: Given $y$, construct the posterior density $p(x|y)$. 

Left: current $I$ and electrode potential $y$; Right: conductivity $x$. 

EIT, Forward/Inverse Problem (image by Siltanen)
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.
Finally, we consider the $\ell_1$ prior case:

$$p(x) \propto \exp \left(-\delta \|Dx\|_1\right),$$

where $D$ is an invertible matrix. Then the posterior then takes the form

$$p(x|y) \propto \exp \left(-\frac{1}{2} \|A(x) - y\|^2 - \delta \|Dx\|_1\right).$$

Note that total variation in one-dimension and the Besov $B^s_{1,1}$-space priors in one- and higher-dimensions have this form.
Finally, we consider the $\ell_1$ prior case:

$$p(x) \propto \exp (-\delta \|Dx\|_1),$$

where $D$ is an invertible matrix. Then the posterior then takes the form

$$p(x|y) \propto \exp \left(-\frac{1}{2}\|A(x) - y\|^2 - \delta \|Dx\|_1\right).$$

Note that total variation in one-dimension and the Besov $B_{1,1}^s$-space priors in one- and higher-dimensions have this form.

* But $p(x|y)$ does not have least squares form.
Prior Transformation for $\ell_1$ Priors

**Main idea:** Transform the problem to one that RTO can solve
- Define a map between a reference parameter $u$ and the physical parameter $x$.
- Choose the mapping so that the prior on $u$ is Gaussian.
- Sample from the transformed posterior, in $u$, using RTO, then transform the samples back.
Prior Transformation for $\ell_1$ Priors

**Main idea:** Transform the problem to one that RTO can solve

- Define a map between a **reference** parameter $u$ and the **physical** parameter $x$.
- Choose the mapping so that the prior on $u$ is Gaussian.
- Sample from the transformed posterior, in $u$, using RTO, then transform the samples back.
The One-Dimensional Transformation

The prior transformation is analytic and is defined

\[ x = S(u) \overset{\text{def}}{=} F_{p(x)}^{-1}(\varphi(u)), \]

where

- \( F_{p(x)}^{-1} \) is the inverse-CDF of the \( L^1 \)-type prior \( p(x) \);
- \( \varphi \) is the CDF of a standard Gaussian.
The One-Dimensional Transformation

The prior transformation is analytic and is defined

\[ x = S(u) \overset{\text{def}}{=} F_{p(x)}^{-1}(\varphi(u)), \]

where

- \( F_{p(x)}^{-1} \) is the inverse-CDF of the \( L^1 \)-type prior \( p(x) \);
- \( \varphi \) is the CDF of a standard Gaussian.

Then the posterior density \( p(x|y) \) can be expressed in terms of \( r \):

\[
p(S(u)|y) \propto \exp \left( -\frac{1}{2} (f(S(u)) - y)^2 - \frac{1}{2} u^2 \right)
= \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} f(S(u)) \\ u \end{bmatrix} - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2 \right)
\]
Prior Transformation: 1D Laplace Prior

\[ p(x) \propto \exp(-\lambda |x|) \]

\[ p(u) \propto \exp\left(-\frac{1}{2}u^2\right) \]

For multiple independent \(x_i\), transformation is repeated
Transformation moves complexity from prior to likelihood
1. Define a change of variables

\[ \mathbf{D} \mathbf{x} = S(\mathbf{u}) \]

such that the transformed prior is a standard Gaussian, i.e.,

\[ p(\mathbf{D}^{-1} S(\mathbf{u})) \propto \exp \left( -\frac{\delta}{2} \| \mathbf{u} \|_2^2 \right) . \]
Laplace Priors in Higher-Dimensions

1. Define a change of variables

\[ \mathbf{Dx} = S(u) \]

such that the transformed prior is a standard Gaussian, i.e.,

\[ p(\mathbf{D}^{-1}S(u)) \propto \exp \left( -\frac{\delta}{2} \| \mathbf{u} \|_2^2 \right). \]

2. Sample from the transformed posterior, with respect to \( \mathbf{u} \),

\[ p(\mathbf{D}^{-1}S(u)|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{D}^{-1}S(u))p(\mathbf{D}^{-1}S(u)); \]
1. Define a change of variables

\[ Dx = S(u) \]

such that the transformed prior is a standard Gaussian, i.e.,

\[ p(D^{-1}S(u)) \propto \exp \left( -\frac{\delta}{2} \|u\|_2^2 \right). \]

2. Sample from the transformed posterior, with respect to \( u \),

\[ p(D^{-1}S(u)|y) \propto p(y|D^{-1}S(u))p(D^{-1}S(u)); \]

3. Transform the samples back via \( x = D^{-1}S(u) \).
Test Case 3, $L^1$-type priors: High-Dimensional Problems

The transformed posterior, with $D$ an invertible matrix, takes the form

$$p(D^{-1}S(u)|y) \propto \exp \left( -\frac{1}{2} \left( f(D^{-1}S(u)) - y \right)^2 - \frac{1}{2}u^2 \right)$$

$$= \exp \left( -\frac{1}{2} \left\| \begin{bmatrix} A(D^{-1}S(u)) \ 0 \end{bmatrix} \right\| - \begin{bmatrix} y \\ 0 \end{bmatrix} \right)^2 \right),$$

where

$$S(u) = (S(u_1), \ldots, S(u_n))$$

as defined above.
Test Case 3, $L^1$-type priors: High-Dimensional Problems

The transformed posterior, with $D$ an invertible matrix, takes the form

$$p(D^{-1}S(u)|y) \propto \exp \left(-\frac{1}{2}(f(D^{-1}S(u)) - y)^2 - \frac{1}{2}u^2\right)$$

$$= \exp \left(-\frac{1}{2} \left\| \begin{bmatrix} A(D^{-1}S(u)) \\ u \end{bmatrix} - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2 \right),$$

where

$$S(u) = (S(u_1), \ldots, S(u_n))$$

as defined above.

$\star p(D^{-1}S(u)|y)$ is in least squares form with respect to $u$ so we can apply RTO!
The RTO Metropolis-Hastings Algorithm

1. Choose $u^0 = u_{\text{MAP}} = \arg\min_u p(D^{-1}S(u)|y)$ and number of samples $N$. Set $k = 1$.
2. Compute an RTO sample $u^* \sim q(u^*)$.
3. Compute the acceptance probability
   \[ r = \min \left( 1, \frac{c(u^{k-1})}{c(u^*)} \right). \]
4. With probability $r$, set $u^k = u^*$, else set $u^k = u^{k-1}$.
5. If $k < N$, set $k = k + 1$ and return to Step 2.
Deconvolution of a Square Pulse w/ TV Prior

\[ p(x|y) \propto \exp \left( -\frac{\lambda}{2} \|Ax - y\|^2 - \delta \|Dx\|_1 \right) \]

\[ x \in \mathbb{R}^{63}, \quad y \in \mathbb{R}^{32} \]
Deconvolution of a Square Pulse w/ TV Prior

![Graphs showing deconvolution results.]

- **RTO**
  - Graph showing deconvolution with RTO algorithm.
  - Axes: $x$ vs. $y$.
  - Annotations: Evaluations.

- **Gibbs**
  - Graph showing deconvolution with Gibbs algorithm.
  - Axes: $x$ vs. $y$.
  - Annotations: Evaluations.

- **MCMC chain**
  - Graph showing MCMC chain for RTO and Gibbs.
  - Axes: Evaluations vs. Chained values.
  - Annotations: $\times 10^6$.
2D elliptic PDE inverse problem

\[-\nabla \cdot (\exp(x(t))\nabla y(t)) = h(t), \quad t \in [0, 1]^2,\]

with boundary conditions

\[\exp(x(t))\nabla y(t) \cdot \vec{n}(t) = 0.\]

After discretization, this defines the model

\[y = A(x).\]
2D PDE inverse problem: mean and STD

Use RTO-MH to sample from the transformed posterior:

\[ p(D^{-1}S(u)|y) \propto \exp \left( -\frac{\lambda}{2} \| A(D^{-1}S(u)) - y \|_2^2 - \delta \| u \|_2^2 \right), \]

where \( D \) is a wavelet transform matrix, then transform the samples back via \( x = D^{-1}S(u) \).
2D PDE inverse problem: Samples
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.
- RTO was described in some detail and then tested on several examples, including EIT and $\ell_1$ priors such as total variation.