Introduction to Bayesian inversion & MCMC

Ville Kolehmainen¹

¹Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

CoE Summer School, Helsinki, June 2019
Contents

Introduction

Bayesian inversion

Gaussian case

About MAP estimate

MCMC

Numerical examples
What are inverse problems?

- Consider measurement problems; estimate the quantity of interest $x \in \mathbb{R}^n$ from (noisy) measurement of $A(x) \in \mathbb{R}^d$ where $A$ is known mapping.

- Inverse problems are characterized as those measurement problems which are *ill-posed*:
  1. The problem is non-unique (e.g., less measurements than unknowns)
  2. Solution is unstable w.r.t measurement noise and modelling errors (e.g., model reduction, inaccurately known nuisance parameters in the model $A(x)$).
An introductory example:

- Consider 2D deconvolution (image deblurring); Given noisy and blurred image

\[ m = Ax + e, \quad m \in \mathbb{R}^n \]

the objective is to reconstruct the original image \( x \in \mathbb{R}^n \).

- Forward model \( x \mapsto Ax \) implements discrete convolution (here the convolution kernel is Gaussian blurring kernel with std of 6 pixels).

Left; Original image \( x \). Right; Observed image \( m = Ax + e \).
Example (cont.)

- Matrix $A$ has trivial null-space $\text{null}(A) = \{0\} \Rightarrow$ solution is unique (i.e. $\exists (A^T A)^{-1}$).
- However, the problem is unstable ($A$ has "unclear" null-space, i.e., $\|A w\| = \| \sum_k \sigma_k \langle w, v_k \rangle u_k \| \approx 0$ for certain $\|w\| = 1$)
- Figure shows the least squares (LS) solution

$$x_{LS} = \arg\min_{x} \| m - A x \|^2 \Rightarrow x_{LS} = (A^T A)^{-1} A^T m$$

Left; true image $x$. Right; LS solution $x_{LS}$
Regularization.

- The ill posed problem is replaced by a well posed approximation. Solution “hopefully” close to the true solution.
- Typically modifications of the associated LS problem
  \[ \| m - Ax \|^2. \]
- Examples of methods; Tikhonov regularization, truncated iterations.
- Consider the LS problem
  \[
  x_{LS} = \arg \min_x \{ \| m - Ax \|^2 \} \Rightarrow (A^T A) x_{LS} = A^T m
  \]

Uniqueness; \( \exists B^{-1} \) if null\( (A) = \{ 0 \} \). Stability of the solution?
Regularization

- Example (cont.); In Tikhonov regularization, the LS problem is replaced with

\[ x_{TIK} = \arg \min_x \{ \| m - Ax \|_2^2 + \alpha \| x \|_2^2 \} \Rightarrow (A^T A + \alpha I) x_{TIK} = A^T m \]

Uniqueness; \( \alpha > 0 \Rightarrow \text{null}(\tilde{B}) = \{0\} \Rightarrow \exists \tilde{B}^{-1} \). Stability of the solution guaranteed by choosing \( \alpha \) “large enough”.

- Regularization poses (implicit) prior about the solution. These assumptions are sometimes well hidden.

Left; true image \( x \). Right; Tikhonov solution \( x_{TIK} \).
Statistical (Bayesian) inversion.

- The inverse problem is recast as a problem of Bayesian inference. The key idea is to extract estimates and assess uncertainty about the unknowns based on:
  - Measured data
  - Model of measurement process
  - Model of a priori information

- Ill-posedness removed by explicit use of prior models!
- Systematic handling of model uncertainties and reduction (⇒ approximation error theory)

Left; true image $x$. Right; Bayesian estimate $x_{CM}$. 
Bayesian inversion.

- All variables in the model are considered as random variables. The randomness reflects our uncertainty of their actual values.
- The degree of uncertainty is coded in the probability distribution models of these variables.
- The complete model of the inverse problem is the posterior probability distribution

\[
\pi(x \mid m) = \frac{\pi_{pr}(x)\pi(m \mid x)}{\pi(m)}, \quad m = m_{\text{observed}}
\]  

where

- \(\pi(m \mid x)\) is the likelihood density; model of the measurement process.
- \(\pi_{pr}(x)\) is the prior density; model for \textit{a priori} information.
- \(\pi(m)\) is a normalization constant.
The posterior density model is a function on $n$ dimensional space;

$$\pi(x \mid m) : \mathbb{R}^n \mapsto \mathbb{R}_+$$

where $n$ is usually large.

To obtain "practical solution", the posterior model is summarized by estimates that answer to questions such as;

- "What is the most probable value of $x$ over $\pi(x \mid m)$?"
- "What is the mean of $x$ over $\pi(x \mid m)$?"
- "In what interval are the values of $x$ with 90% (posterior) probability?"
Point estimates

- Maximum a posteriori (MAP) estimate:
  \[ \pi(x_{MAP} \mid m) = \arg \max_{x \in \mathbb{R}^n} \pi(x \mid m). \]

- Conditional mean (CM) estimate:
  \[ x_{CM} = \int_{\mathbb{R}^n} x \pi(x \mid m) \, dx. \]
Maximum a posteriori (MAP) estimate:
\[ \text{arg max}_{x \in \mathbb{R}^n} \pi(x \mid m) \]

Conditional mean (CM) estimate:
\[ \int_{\mathbb{R}^n} x \pi(x \mid m) \, dx \]
Spread estimates

- Covariance:

\[ \Gamma_{x|m} = \int_{\mathbb{R}^n} (x - x_{CM})(x - x_{CM})^T \pi(x \mid m) \, dx \]

- Confidence intervals; Given \(0 < \tau < 100\), compute \(a_k\) and \(b_k\) s.t.

\[ \int_{-\infty}^{a_k} \pi_k(x_k) \, dx_k = \int_{b_k}^{\infty} \pi_k(x_k) \, dx_k = \frac{100 - \tau}{200} \]

where \(\pi_k(x_k)\) is the marginal density

\[ \pi_k(x_k) = \int_{\mathbb{R}^{n-1}} \pi(x \mid m) \, dx_1 \cdots dx_{k-1} \, dx_{k+1} \cdots dx_n. \]

The interval \(I_k(\tau) = [a_k \ b_k]\) contains \(\tau\%\) of the mass of the marginal density.
Illustration of confidence intervals

- \( x = (x_1, x_2)^T \in \mathbb{R}^2 \)
- \( \pi(x_1|m) = \int \pi(x|m)dx_2 \) (middle)
- \( \pi(x_2|m) = \int \pi(x|m)dx_1 \) (right)
- Solid vertical lines show the mean value \( x_{CM} \), dotted the 90% confidence limits.

Left; Contours of \( \pi(x|m) \). \( x_{CM} \) is shown with +. Middle; marginal density \( \pi(x_1|m) \). Right; marginal density \( \pi(x_2|m) \).
Gaussian $\pi(x|m)$:

- Let $m = Ax + e$, $e \sim \mathcal{N}(e_*, \Gamma_e)$ independent of $x$ and $\pi_{pr}(x) = (x_*, \Gamma_x) \rightarrow$:

\[ \pi(x | m) \propto \pi(m | x)\pi_{pr}(x) \]
\[ = \exp \left( -\frac{1}{2}(m - Ax - e_*)^T\Gamma_e^{-1}(m - Ax - e_*) \right) \]
\[ - \frac{1}{2}(x - x_*)^T\Gamma_x^{-1}(x - x_*) \]

- $\pi(x | m)$ fully determined by mean and covariance:

\[ x_{CM} = \Gamma_{x|m}(A^T\Gamma_e^{-1}(m - e_*) + \Gamma_x^{-1}x_*) \]
\[ \Gamma_{x|m} = (A^T\Gamma_e^{-1}A + \Gamma_x^{-1})^{-1} \]
Let
\[ L_e^T L_e = \Gamma_e^{-1}, \quad L_x^T L_x = \Gamma_x^{-1}, \]
then we can write \( \pi(x \mid m) \propto \exp\{-\frac{1}{2} F(x)\} \), where
\[
F(x) = \|L_e(m - Ax - e_*)\|_2^2 + \|L_x(x - x_*)\|_2^2
\]
and
\[
x_{\text{MAP}} = \arg\min_x F(x)
\]
⇒ Connection to Tikhonov regularization!
Example

- Consider the original form of Tikhonov regularization:

\[ x_{\text{TIK}} = \arg \min_x \{ \| m - Ax \|_2^2 + \alpha \| x \|_2^2 \} \quad \Rightarrow \quad x_{\text{TIK}} = (A^T A + \alpha I)^{-1} A^T m \]

- From the Bayesian viewpoint, \( x_{\text{TIK}} \) correspond to \( x_{\text{CM}} \) with the following assumptions:
  - Measurement model \( m = Ax + e \), \( x \) and \( e \) mutually independent with \( e \sim \mathcal{N}(0, I) \).
  - \( x \) is assumed \textit{a priori} mutually independent zero mean white noise with variance \( 1/\alpha \).

- The original idea of the Tikhonov method was to approximate \( A^T A \) with a matrix \( A^T A + \alpha I \) that is invertible and produces stable solution.
About computation of MAP estimate

- Solving MAP estimate

\[ x_{\text{MAP}} = \arg \max_x \pi(x \mid m) \]

is an optimization problem.

- Example;

\[ \pi(x \mid m) \propto \pi_{+}(x) \exp \left( -\frac{1}{2} \| L_e(m - A(x)) \|_2^2 - W(x) \right) \]

\[ \Rightarrow \]

\[ x_{\text{MAP}} = \arg \min_{x \geq 0} \left\{ \frac{1}{2} \| L_e(m - A(x)) \|_2^2 + W(x) \right\} \]

- Large variety of optimization methods
Computation of the integration based estimates

- Many estimators are of the form

\[ \bar{f}(x) = \int_{\mathbb{R}^n} f(x) \pi(x \mid m) \, dx \]

- Examples:
  - \( f(x) = x \leadsto x_{CM} \)
  - \( f(x) = (x - x_{CM})(x - x_{CM})^T \leadsto \Gamma_x \mid m \)
  - etc ...

- Analytical evaluation in most cases impossible
- Traditional numerical quadratures not applicable when \( n \) is large (number of points needed unreasonably large, support of \( \pi(x \mid m) \) may not be well known)
- \Rightarrow Monte Carlo integration.
Monte Carlo integration

- Monte Carlo integration
  1. Draw an ensemble \( \{x^{(k)}, k = 1, 2, \ldots, N\} \) of i.i.d samples from \( \pi(x) \).
  2. Estimate

\[
\int_{\mathbb{R}^n} f(x) \pi(x) dx \approx \frac{1}{N} \sum_{k=1}^{N} f(x^{(k)})
\]

- Convergence (law of large numbers)

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(x^{(k)}) \to \int_{\mathbb{R}^n} f(x) \pi(x) dx \quad \text{(a.c)}
\]

- Variance of the estimator \( \bar{f} = \frac{1}{N} \sum_{k=1}^{N} f(x^{(k)}) \) reduces

\[
\propto \frac{1}{N}
\]
Simple example of Monte Carlo integration

1. Let \( x \in \mathbb{R}^2 \) and

\[
\pi(x) = \begin{cases} 
\frac{1}{4}, & x \in G, \\
0, & x \not\in G.
\end{cases}
\]

\[ G = [-1, 1] \times [-1, 1] \]

The task is to compute integral \( g(x) = \int_{\mathbb{R}^2} \chi_D \pi(x) \, dx \) where \( D \) is the unit disk on \( \mathbb{R}^2 \).

2. Using \( N = 5000 \) samples, we get estimate \( g(x) = 0.7814 \) (true value \( \pi/4 = 0.7854 \))
MCMC

- Direct sampling of the posterior usually not possible ⇒ Markov chain Monte Carlo (MCMC).
- MCMC is Monte Carlo integration using Markov chains (dependent samples);
  1. Draw \( \{x^{(k)}, k = 1, 2, \ldots, N\} \sim \pi(x) \) by simulating a Markov chain (with equilibrium distribution \( \pi(x) \)).
  2. Estimate

\[
\int_{\mathbb{R}^n} f(x) \pi(x) dx \approx \frac{1}{N} \sum_{k=1}^{N} f(x^{(k)})
\]

- Variance of the estimator reduces as \( \propto \tau/N \) where \( \tau \) is the integrated autocorrelation time of the chain.
- Algorithms for MCMC;
  1. Metropolis Hastings algorithm
  2. Gibbs sampler
Metropolis-Hastings algorithm

- Generation of the ensemble \( \{x^{(k)}, k = 1, \ldots, N\} \sim \pi(x) \) using Metropolis Hastings algorithm;
  1. Pick an initial value \( x^{(1)} \in \mathbb{R}^n \) and set \( \ell = 1 \)
  2. Set \( x = x^{(\ell)} \).
  3. Draw a candidate sample \( x' \) from proposal density
     \[ x' \sim q(x, x') \]
     and compute the acceptance factor
     \[ \alpha(x, x') = \min \left( 1, \frac{\pi(x')q(x', x)}{\pi(x)q(x, x')} \right). \]
  4. Draw \( t \in [0, 1] \) from uniform probability density \( t \sim \text{uni}(0, 1) \).
  5. If \( \alpha(x, x') \geq t \), set \( x^{(\ell+1)} = x' \), else \( x^{(\ell+1)} = x \). Increment \( \ell \rightarrow \ell + 1 \).
  6. When \( \ell = N \) stop, else repeat from step 2.
Normalization constant of $\pi$ does not need to be known.

Great flexibility in choosing the proposal density $q(x, x')$; almost any density would do the job (eventually).

However, the choice of $q(x, x')$ is a crucial part of successful MCMC; it determines the efficiency (autocorrelation time $\tau$) of the algorithm $\Rightarrow q(x, x')$ should be s.t. $\tau$ as small as possible.

No systematic methods for choosing $q(x, x')$. More based on “intuition & art”.
The updating may proceed:
- All unknowns at a time
- Single component $x_j$ at a time
- A block of unknowns at a time

The order of updating (in single component and blockwise updating) may be:
- Update elements chosen in random (random scan)
- Systematic order

The proposal can be a mixture of several densities:

$$q(x, x') = \sum_{i=1}^{t} p_i q_i(x, x'), \quad \sum_{i=1}^{t} p_i = 1$$
Parameters of $q(x, x')$ are usually calibrated by pilot runs; aim at finding parameters giving best efficiency (minimal $\tau$).

Determining the “burn-in”; Trial runs (e.g, running several chains & checking that they are consistent). Often determined from the plot of “posterior trace” \( \{\pi(x^{(k)}), k = 1, \ldots, N\} \) (see Figure).
Example (Metropolis-Hastings)

Let $x \in \mathbb{R}^2$ and posterior

$$\pi \propto \pi_D(x) \exp \left\{ -10(x_1^2 - x_2)^2 - (x_2 - \frac{1}{4})^4 \right\},$$

where $\pi_D$ is

$$\pi_D(x) = \begin{cases} 1, & x \in D \\ 0, & \text{otherwise} \end{cases}$$

$$D = [-2, 2] \times [-2, 2] \subset \mathbb{R}^2.$$
Sample distribution $\pi(x)$ using the Metropolis Hastings algorithm. Draw proposal $x' \in \mathbb{R}^2$ as

$$x'_i = x_i + \gamma \xi_i, \quad \xi_i \sim \mathcal{N}(0, 1). \quad i = 1, 2$$

$$\Rightarrow$$

$$q(x, x') = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi\gamma}} \exp \left\{ -\frac{1}{2\gamma^2} (x'_i - x_i)^2 \right\}.$$ 

Consider the effect of the parameter $\gamma$ on the efficiency of the algorithm.
Sampling with $\gamma = 0.02$

$\tau = 187.3$, acceptance ratio 95%.

Chain moves too slowly (very small steps $\rightarrow$ long correlation), inefficient.

Left; samples. Middle; Posterior trace $\{\pi(x^{(k)})\}$ (400 states). Right; scaled autocovariance function.
Sampling with $\gamma = 0.4$

$\tau = 7.4$, acceptance ratio 40%.

Chain is moving optimally.

Left; samples. Middle; Posterior trace $\{\pi(x^{(k)})\}$ (400 states). Right; scaled autocovariance function.
Sampling with $\gamma = 1.8$

- $\tau = 33.8$, acceptance ratio 7%.
- Proposal too wide; most proposals $x'$ rejected, chain gets "stuck" to single state for long periods ($\rightarrow$ long correlation).

Left; samples. Middle; Posterior trace $\{\pi(x^{(k)})\}$ (400 states). Right; scaled autocovariance function.
Summary of the example; Left image shows $\tau$ vs $\gamma$ and right image shows acceptance ratio vs. $\gamma$.

”Rule of thumb” (when $q(x, x')$ Gaussian); Aim at acceptance ratio 20 – 50%.
Example (cont.)

- Improve the proposal. Let us draw samples as

\[ x'_i = x_i + \delta_i(x) + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \gamma^2), \ i = 1, 2, \]

where \( \delta_i : \mathbb{R}^2 \mapsto \mathbb{R} \) is deterministic mapping. Now

\[
q(x, x') = \prod_{i=1}^{2} \frac{1}{\sqrt{2\pi\gamma}} \exp \left\{ -\frac{1}{2\gamma^2} (x'_i - (x_i + \delta_i(x))^2) \right\}.
\]

- Choose \( \delta_i(x) = (h\nabla \pi(x))_i \), where \( h \) is constant.
- Sampling with $\gamma = 0.4$, $h = 0.02$
- $\tau = 6.9$, acceptance ratio 44%.

Left; samples. Middle; Posterior trace $\{\pi(x^{(k)})\}$ (400 states). Right; scaled autocovariance function.
Gibbs sampler

- Generation of the ensemble \( \{x^{(k)}, k = 1, \ldots, N\} \sim \pi(x) \) using Gibbs sampler:
  1. Pick initial value \( x^{(0)} \) and set \( j = 1 \).
  2. Generate \( x^{(j)} \) a single variable at a time:

        draw \( x_1^{(j)} \) from the density \( t \mapsto \pi(t|x_2^{(j-1)}, \ldots, x_n^{(j-1)}) \),
        draw \( x_2^{(j)} \) from the density \( t \mapsto \pi(t|x_1^{(j)}, x_3^{(j-1)}, \ldots, x_n^{(j-1)}) \),
                \vdots
        draw \( x_n^{(j)} \) from the density \( t \mapsto \pi(t|x_1^{(j)}, \ldots, x_{n-1}^{(j)}) \).

  3. If \( j = N \), stop, else set \( j \leftarrow j + 1 \) and go to 2.
Illustration of the componentwise sampling with the Gibbs sampler.

- Conditionals can be sampled as follows;
  1. Determine the cumulative function

\[
\Phi_j(t) = \int_{-\infty}^{t} \pi(x_j | x_{-j}) dx_j,
\]

where \( x_{-j} = (x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) \in \mathbb{R}^{n-1} \)

2. Draw random sample \( \xi \sim \text{uni}([0 1]) \). Sample \( y_j \) is obtained as

\[ y_j = \Phi_j^{-1}(\xi) \]
Illustration of sampling from the 1D conditionals.

- Top: $\pi(x_j \mid x_{-j})$
- Bottom: cumulative function $\Phi_j(t) = \int_{-\infty}^{t} \pi(x_j \mid x_{-j})dx_j$.
- $\xi \sim \text{uni}(0, 1)$ (location of horizontal line)
- Sample $y_j = \Phi_j^{-1}(\xi)$ (location of the vertical line)
Posterior typically in non-parametric form, normalization constant unknown ⇒ Numerical approximation (e.g. trapezoidal quadrature)

\[
\Phi_j(t_m) = C \int_a^{t_m} \pi(x_j \mid x_{-j})dx_j \approx C \sum_{k=1}^{m} w_k \pi(t_k \mid x_{-j})
\]

where \(w_k\) are the quadrature weights.

Support \([a, b]\) of the conditional \(\pi(x_j \mid x_{-j})\) has to be “forked” carefully. \(C\) chosen s.t. \(\Phi(b) = 1\).

Sample can be obtained by interpolation. Let \(\Phi_j(t_p) < \xi < \Phi_j(t_{p+1})\), then linear interpolation gives

\[
y_j = t_p + \frac{\xi - \Phi_j(t_p)}{\Phi_j(t_{p+1}) - \Phi_j(t_p)}(t_{p+1} - t_p)
\]
Some features of Gibbs sampler;

- Acceptance probability always 1. No need to tuning the proposal.
- Determination of burn-in similarly as for Metropolis Hastings.
- If $\pi(x)$ has high correlation, can get inefficient ($\tau$ increases)
- Numerical evaluation of the conditionals can get computationally infeasible (e.g. high dimensional cases with PDE based forward models).
Example (cont.); Gibbs sampling

- $\tau = 1.9$.
- Here computation time longer than for Metropolis Hastings.

Left; samples. Middle; Posterior trace $\{\pi(x^{(k)})\}$ (400 states). Right; scaled autocovariance function.
Example; Estimates (Gibbs sampling)

- \( x_{CM} \approx (0.00, 0.35)^T \)
- Marginal densities \( \pi(x_1) = \int \pi(x)dx_2 \) and \( \pi(x_2) = \int \pi(x)dx_1 \)
- Solid vertical lines show the mean, dotted the 90% confidence limits.

Left; Contours of \( \pi(x) \). \( x_{CM} \) is shown with +. Middle; marginal density \( \pi(x_1) \). Right; marginal density \( \pi(x_2) \).
Numerical examples

In summary, Bayesian solution of an inverse problem consist of the following steps:

1. Construct the likelihood model $\pi(m|x)$. This step includes the development of the model $x \mapsto A(x)$ and modeling the measurement noise and modelling errors.
2. Construction of the prior model $\pi_{pr}(x)$.

We consider the following numerical "classroom" examples:

1. 2D deconvolution (intro example continued)
2. Limited data emission tomography
3. EIT with experimental data
Example 1; 2D deconvolution

- We are given a blurred and noisy image

\[ m = Ax + e, \quad m \in \mathbb{R}^n \]

- Assume that we know \textit{a priori} that the true solution is binary image representing some text.

Left; true image \emph{x}. Right; Noisy, blurred image \( m = Ax + e \).
Likelihood model $\pi(m \mid x)$:

- Joint density
  \[ \pi(m, x, e) = \pi(m \mid x, e) \pi(e \mid x) \pi(x) = \pi(m, e \mid x) \pi(x) \]

- In case of $m = Ax + e$, we have
  \[ \pi(m \mid x, e) = \delta(m - Ax - e), \text{ and} \]

\[
\begin{align*}
\pi(m \mid x) & = \int \pi(m, e \mid x) \, de \\
& = \int \delta(m - Ax - e) \pi(e \mid x) \, de \\
& = \pi_{e \mid x}(m - Ax \mid x)
\end{align*}
\]
In the (usual) case of mutually independent $x$ and $e$, we have $\pi_{e|x}(e|x) = \pi_e(e)$ and

$$\pi(m|x) = \pi_e(m - Ax)$$

The model of the noise: $\pi(e) = \mathcal{N}(0, \Gamma_e)$ →,

$$\pi(m|x) \propto \exp \left( -\frac{1}{2} \left( \| L_e(m - Ax) \|^2 \right) \right),$$

where $L_e^T L_e = \Gamma_e^{-1}$.

Remark; we have assumed that the (numerical) model $Ax$ is exact (i.e., no modelling errors).
Prior model $\pi_{pr}(x)$

- The prior knowledge:
  - $x$ can obtain the values $x_j \in \{0, 1\}$
  - Case $x_j = 0$ (black, background), case $x_j = 1$ (white, text).
  - The pixels with value $x_j = 1$ are known to be clustered in "blocky structures" which have short boundaries.

- We model the prior knowledge with Ising prior

$$
\pi_{pr}(x) \propto \exp \left( \alpha \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} \delta(x_i, x_j) \right)
$$

where $\delta$ is the Kronecker delta

$$
\delta(x_i, x_j) = \begin{cases} 
1, & x_i = x_j \\
0, & x_i \neq x_j 
\end{cases}
$$
Posterior model $\pi(x \mid m)$:

- Posterior model for the deblurring problem;

$$
\pi(x|m) \propto \exp \left( -\frac{1}{2} \| L_e (m - Ax) \|_2^2 + \alpha \sum_{i=1}^{n} \sum_{j \in \mathcal{N}_i} \delta(x_i, x_j) \right)
$$

- Posterior explored with the Metropolis Hastings algorithm.
Metropolis Hastings proposal

- The proposal is a mixture $q(x, x') = \sum_{i=1}^{2} \xi_i q_i(x, x')$ of two move types.

  **Move 1:** choose update element $x_i \in \mathbb{R}$ by drawing index $i$ with uniform probability $\frac{1}{n}$ and change the value of $x_i$.

  **Move 2:** Let $N^*(x)$ denote the set of active edges in image $x$ (edge $l_{ij}$ connects pixels $x_i$ and $x_j$ in the lattice; it is active if $x_i \neq x_j$). Pick an active update edge w.p. $\frac{1}{|N^*(x)|}$

  Pick one of the two pixels related to the chosen edge w.p. $\frac{1}{2}$ and change the value of the chosen pixel.
Results

- Left image; True image $x$
- Middle; Tikhonov regularized solution
  \[ x_{\text{TIK}} = \arg\min_x \{ \| m - Ax \|_2^2 + \alpha \| x \|_2^2 \} \]
- Right image; CM estimate $x_{\text{CM}} = \int_{\mathbb{R}^n} x \pi(x|m)dx$.

Left; true image $x$, Middle; $x_{\text{TIK}}$. Right; $x_{\text{CM}}$. 
Results

- Left image; $x_{CM}$
- Posterior variances $\text{diag}(\Gamma x|m)$
- Right; A sample from $\pi(x \mid m)$
Example 2; Emission tomography for brachytherapy

- Brachytherapy (sealed source radiotherapy); radioactive source pellets are placed inside the tumor using a needle and pneumatic loader.
- Used in treatment of cancers of prostate, breast, head and neck area, etc...
TLD emission data (six 1D projections with angular separation of 30 degrees).

- Objective; use (limited data) emission tomography for verification of the correct placement of the source pellets inside the tissues.

- Phantom experiment. 6 projections with projection interval of 30°

- Data collected with a termoluminescent dosimeter (TLD) with a parallel beam collimator geometry. 41 readings in each projection (i.e., \( m \in \mathbb{R}^{246} \))
Forward model

- The model for the expectation of the observed photon count

\[ \bar{m}_j = \int_{L_j} x(s) ds \]

where \( L_j \) is the “line of sight” detected from the \( j \):th detector position.

- The model neglects scattering phenomena and attenuation correction.

- Discretization; the domain \( \Omega \) of interest is divided to regular \( 41 \times 41 \) pixel grid (i.e., \( f \in \mathbb{R}^{1681} \)).

- The forward model becomes

\[ x \mapsto Ax, \quad A : \mathbb{R}^{1681} \mapsto \mathbb{R}^{246} \]

- The inverse problem is to estimate the activity distribution \( x \), given the vector of observed photon counts \( m \).
Likelihood $\pi(m|x)$:

- Likelihood model; the observations $\{m_j, j = 1, \ldots, n_m\}$ are Poisson distributed random variables. The fluctuations are assumed mutually independent $\Rightarrow$

$$
\pi(m \mid x) = \prod_{j=1}^{n_m} \frac{(Ax)_j^{m_j}}{m_j!} \exp(-(Ax)_j)
\propto \exp(m^T \log(Ax) - 1^T(Ax))
$$

- The model neglects electronic noise of the data acquisition system.
Prior model $\pi_{pr}(x)$:

- We know \textit{a priori} that the activity distribution is:
  - Non-negative
  - The activity is contained in small granules (small pixel clusters of approximately constant activity).

We model this knowledge by prior model

$$
\pi_{pr}(x) \propto \pi_+(x) \exp \left( -\alpha \sum_{k=1}^{n} \sum_{j \in \mathcal{N}_k} |x_k - x_j|^p \right), \ p = 1
$$

Images having total variations (from left to right) 18, 28 and 40.
Posterior density;

\[ \pi(x|m) \propto \pi_+(x) \exp(m^T \log(Ax) - 1^T(Ax) - \alpha \sum_{k=1}^{n} \sum_{j \in \mathcal{N}_k} |x_k - x_j|) \]

We compute CM estimate by Metropolis-Hastings MCMC.

Proposals \( x' \) are drawn with the following single component random scan;

**Step 1:** choose update element \( x_i \in \mathbb{R} \) by drawing index \( i \) with uniform probability \( \frac{1}{n} \)

**Step 2:** Generate \( x' \) by updating \( x_i \) s.t.:

\[ x'_i = |x_i + \xi|, \quad \xi \sim \mathcal{N}(0, \epsilon^2) \]
Left image; Tikhonov regularized solution

\[ x_{\text{TIK}} = \arg \min_x \{ \| m - Ax \|^2_2 + \alpha \| x \|^2_2 \} \]

Right image; CM estimate \( x_{\text{CM}} = \int_{\mathbb{R}^n} x \pi(x|m) dx \).

The source pellets (3 pcs.) are localized correctly in the CM estimate.
- Left image; CM estimate $x_{CM}$.
- Right image; posterior variances $\text{diag} \Gamma_{x|m}$
Example 2: "tank EIT"

- Joint work with Geoff Nicholls (Oxford) & Colin Fox (U. Otago, NZ)
- Data measured with the Kuopio EIT device.
- 16 electrodes, adjacent current patterns.
- Target: plastic cylinders in salt water.
Measurement model

\[ V = U(\sigma) + e, \quad e \sim \mathcal{N}(e_*, \Gamma_e) \]

where \( V \) denotes the measured voltages and \( U(\sigma) \) FEM-based forward map.

- \( e \) and \( \sigma \) modelled mutually independent.
- \( e_* \) and \( \Gamma_e \) estimated by repeated measurements.

Posterior model

\[
\pi(\sigma | V) \sim \exp \left\{ -\frac{1}{2} (V - U(\sigma))^T \Gamma_e^{-1} (V - U(\sigma)) \right\} \pi_{pr}(\sigma)
\]

We compute results with three different prior models \( \pi_{pr}(\sigma) \).

Sampling carried out with the Metropolis-Hastings sampling algorithm.
Prior models $\pi_{pr}(\sigma)$

1. **SMOOTHNESS PRIOR AND POSITIVITY PRIOR**

\[
\pi_{pr}(\sigma) \propto \pi_+(\sigma) \exp \left( -\alpha \sum_{i=1}^{n} H_i(\sigma) \right), \quad H_i(\sigma) = \sum_{j \in \mathcal{N}_i} |\sigma_i - \sigma_j|^2.
\]

2. **MATERIAL TYPE PRIOR (FOX & NICHOLLS, 1998)**

- The possible material types \(\{1, 2, \ldots, C\}\) inside the body are known but the segmentation to these materials is unknown. The material distribution is represented by an (auxiliary) discrete valued random field $\tau$ (pixel values $\tau_j \in \{1, 2, \ldots, C\}$).
- The possible values of conductivity $\sigma(\tau)$ for different materials are known only approximately. Gaussian prior for the material conductivities.
- The different material types are assumed to be clustered in “blocky structures” (e.g., organs, etc).
\[ \pi_{pr}(\sigma) = \pi_{pr}(\sigma|\tau)\pi_{pr}(\tau), \]

where

\[ \pi_{pr}(\sigma|\tau) \propto \pi_+(\sigma) \exp\left(-\alpha \sum_{i=1}^{n} G_i(\sigma)\right) \prod_{i=1}^{n} \tilde{\pi}(\sigma_i|\tau_i) \]

where \( \tilde{\pi}(\sigma_i|\tau_i) = \mathcal{N}(\eta(\tau_i), \xi(\tau_i)^2) \)

and

\[ G_i(\sigma) = \sum_{i} \sum_{j \in \{k|k \in \mathcal{N}_i \text{ and } \tau_k = \tau_i\}} (\sigma_i - \sigma_j)^2 \]

\( \pi_{pr}(\tau) \) is an Ising prior:

\[ \pi_{pr}(\tau) \propto \exp \left( \beta \sum_{i=1}^{n} T_i(\tau) \right), \quad T_i(\tau) = \sum_{j \in \mathcal{N}_i} \delta(\tau_i, \tau_j), \quad (2) \]

where \( \delta(\tau_i, \tau_j) \) is the Kronecker delta function.
3 CIRCLE PRIOR

- Domain $\Omega$ with known (constant) background conductivity $\sigma_{bg}$ is assumed to contain unknown number of circular inclusions (i.e., dimension of state not fixed)
- The inclusions have known contrast
- Sizes and locations of inclusions unknown

- For the number $N$ of the circular inclusions we write a point process prior:

$$\pi_{pr}(\sigma) = \beta^N \exp(-\beta A) \delta(\sigma \in A),$$

where

- $\beta = 1/A$ ($A$ is the area of the domain $\Omega$)
- $A$ is set of feasible circle configurations (the circle inclusions are disjoint) and $\delta$ is the indicator function.
CM estimates with the different priors

- Smoothness prior (Top right), material type prior (Bottom left) and circle prior (Bottom right)