The University of Auckland – Applied Mathematics

Computing Sample-Based Solutions to Inverse Problems

Colin Fox

Andrés Christen (CIMAT, Mexico)

fox@math.auckland.ac.nz

Outline

- Naive MCMC: components of straightforward implementation
- Two-step modification to M-H MCMC
- Conjugate direction sampler
- Wrap up

Metropolis-Hastings algorithm

To sample from $\pi(\cdot)$

1. given state x_t at time t generate candidate state x' from a proposal distribution $q\left(.|x_t\right)$
2. With probability $\alpha\left(x_t \rightarrow x'\right) = \min\left(1, \dfrac{\pi(x')q\left(x_t|x'\right)}{\pi(x_t)q\left(x'|x_t\right)}\right)$ set $X_{t+1} = x'$

   otherwise $X_{t+1} = x_t$
3. Repeat

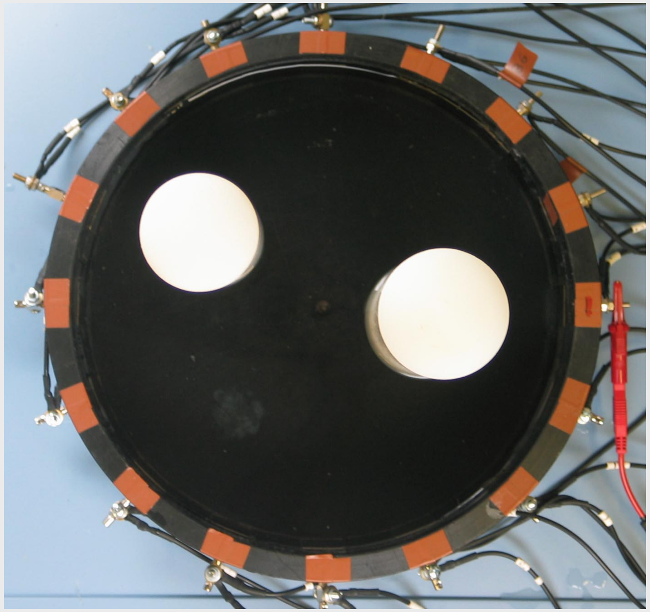
$q\left(.|x_t\right)$ can be any distribution that ensures the chain is irreducible and aperiodic.

Electrical Impedance Tomography

For fixed current patterns $\{I\}$

$$A : \sigma \mapsto \{U\}$$

Simulate A by solving the BVP



$$\nabla \cdot \sigma \nabla u = 0$$

$$\int_{e_l} \sigma \frac{\partial u}{\partial n} dS = I_l$$

$$\sigma \frac{\partial u}{\partial n}\bigg|_{\partial\Omega \setminus \bigcup_l e_l} = 0$$

$$\left(u + z_l \sigma \frac{\partial u}{\partial n} \right)\bigg|_{e_l} = U_l$$

Posterior density

$$\pi(\sigma \mid V) \sim \exp\left\{ - \left(\frac{1}{2} (V - U(\sigma))^{\mathrm{T}} C_n^{-1} (V - U(\sigma)) \right) \right\} \pi_{\mathrm{pr}}(\sigma)$$

Material type prior

$$\pi_{\mathrm{pr}}(\sigma|\tau) \propto \pi_+(\sigma)\exp\left(-\alpha\sum_i \sum_{j\sim i \mathrm{and} \tau_j=\tau_i}(\sigma_i-\sigma_j)^2\right)\prod_{i=1}^{M}\exp\Big(-\frac{1}{2\xi_i^2}(\sigma_i-\eta(\tau_i))^2\Big)$$

Segmented MRF conditioned on material type $\tau_i \in \{1, 2, \ldots, C\}$ modelled as a Potts MRF on the pixel latice

$$\pi_{\mathrm{pr}}(\tau) \propto \exp\left(\beta\sum_{i=1}^{M}\sum_{j\sim i}\delta_{\tau_i,\tau_j}\right)$$

Multiple move M-H algorithm: pick at random

1. Set a random pixel to random type and value
2. Flip a pixel type and value near a material boundary
3. Swap a pair of pixels connected by an update edge
4. Assign new values for pixels connected by an update edge
5. Draw new conductivity value for a pixel

Nicholls and F 1998

Material type prior - results

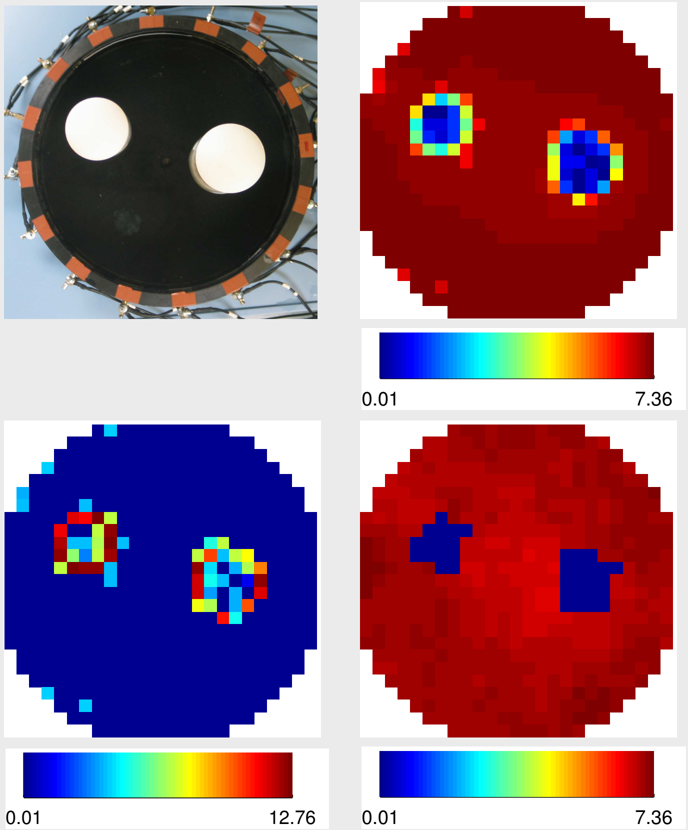


Figure 3: Results with the Material type MRF-prior. Top left: Photograph of the measurement setup. Top right: Posterior mean for the conductivity. Bottom left: Posterior variance of the conductivity. Bottom right: One sample from the posterior.

Kolehmainen, Fox and Nicholls, *MCMC Inversion of Measured EIT Data*, 200?

Speeding-Up MCMC Sampling from $\pi(\cdot)$

1. At $x^{(t)}$ generate proposal y from $q(\cdot \mid x^{(t)})$

2. Let

$$g(x, y) = \min\left\{1, \frac{q(x \mid y)}{q(y \mid x)} \frac{\pi_x^*(y)}{\pi_x^*(x)}\right\}$$

   w.p. $g(x^{(t)}, y)$, "promote" y. New proposal distribution is

$$q^*(y \mid x) = g(x, y) q(y \mid x) + (1 - r(x))\delta_x(y)$$

3. Let

$$\rho(x, y) = \min\left\{1, \frac{q^*(x \mid y)}{q^*(y \mid x)} \frac{\pi(y)}{\pi(x)}\right\}$$

   w.p. $\rho(x^{(t)}, y)$ accept y setting $x^{(t+1)} = y$, otherwise $x^{(t+1)} = x^{(t)}$

Christen F, *MCMC using an Approximation*, Journal of Computational and Graphical Statistics, 2005

Speeding-Up MCMC Sampling from $\pi(\cdot)$

1. At $x^{(t)}$ generate proposal y from $q(\cdot \mid x^{(t)})$

2. Let

   $$g(x, y) = \min\left\{1, \frac{q(x \mid y)}{q(y \mid x)} \frac{\pi_x^*(y)}{\pi_x^*(x)}\right\}$$

   w.p. $g(x^{(t)}, y)$, "promote" y. New proposal distribution is

   $$q^*(y \mid x) = g(x, y)q(y \mid x) + (1 - r(x))\delta_x(y)$$

3. Let

   $$\rho(x, y) = \min\left\{1, \frac{q^*(x \mid y)}{q^*(y \mid x)} \frac{\pi(y)}{\pi(x)}\right\}$$

   w.p. $\rho(x^{(t)}, y)$ accept y setting $x^{(t+1)} = y$, otherwise $x^{(t+1)} = x^{(t)}$

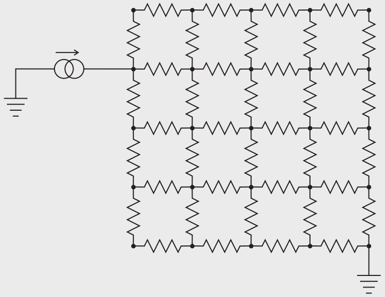
e.g.

$$\pi_x^*(x + \Delta x|d) \propto \exp\left\{-\chi\left(d - (A(x) + J\Delta x)\right) - \rho\left(x\right)\right\}$$

Christen F, *MCMC using an Approximation*, Journal of Computational and Graphical Statistics, 2005

Recovering Resistor Values in a Network

Resistor values, known to be either 2Ω or 3Ω in a square grid with 24 resistors per side, hence $n = 1200$ resistors. Measurements made by injecting current into each of 16 boundary electrodes in turn, and measuring the resulting voltage at all electrodes: 300 independent measurements.



$$(Y_\sigma)_{lm} = \begin{cases} -\sigma_{lm} & l \neq m \\ \displaystyle\sum_{k=1}^{N} \sigma_{lk} & l = m \end{cases}$$

$$Y_\sigma v = i$$

Measurements are of $v_j : j \in E = 1, 2, \ldots, |E| \leq N$ for a set of fixed $i : i_j, j \in E$ being determined, with other components being zero.

Defines forward map

$$A : \sigma \mapsto \left(Y_\sigma^{-1}\right)_{E,E}$$

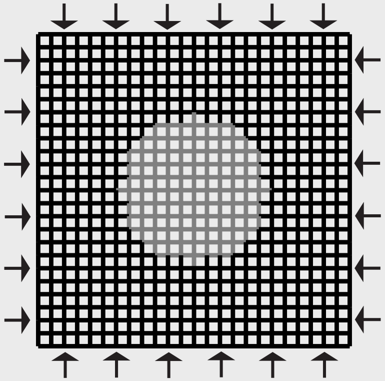
Inverse problem is to find σ from noisy measurement of $\left(Y_\sigma^{-1}\right)_{E,E}$

Approximation and Phantom

- Exact solution: Cholesky factorize Y and solve
- First order approximation: For $\sigma_{lm} \to \sigma_{lm} + \Delta\sigma$ is

$$Y_{EE}^{-1} \to Y_{EE}^{-1} - \Delta\sigma U U^{\mathsf{T}}$$

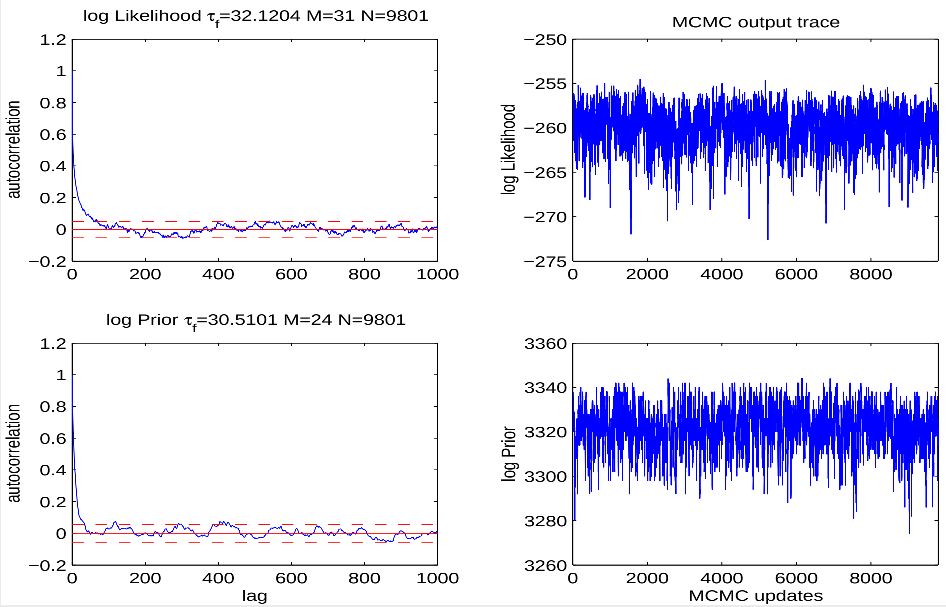
where $U = Y_{El}^{-1} - Y_{Em}^{-1}$ with the associated term absent if $l = N$ or $m = N$.

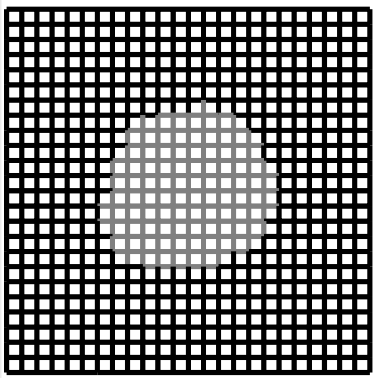


$$\pi(r|D) \propto \exp\left\{ -\frac{\left\| D - Y_{EE}^{-1} \right\|_{\mathsf{F}}^{2}}{2s^{2}} + \theta \sum_{r_{lm} \in r} \sum_{r_{lm} \sim r_{pq}} \delta_{r_{lm}, r_{pq}} \right\}$$

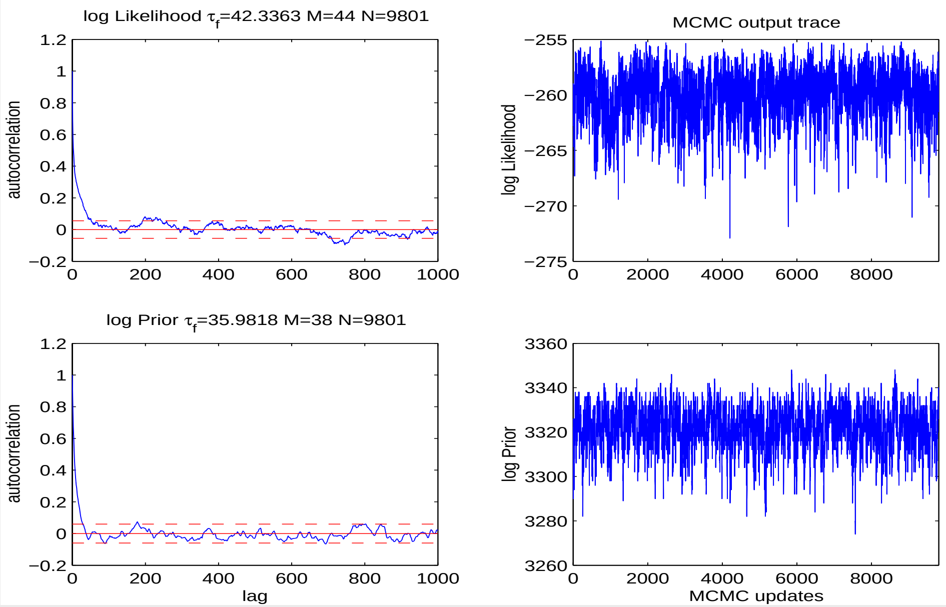
Computational Results

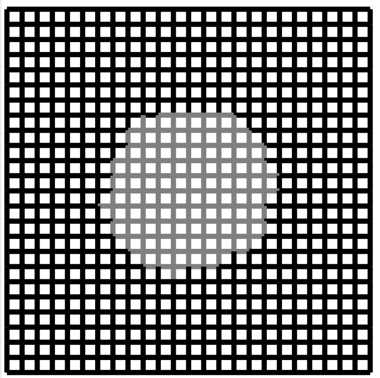
exact





linearization plus correction





Two-step algorithm gives speed up by factor of 25

Time to converge (in expectation) in EIT problem

Using calibrated forward map, complete electrode model, Gaussian smoothness prior in 2-dimensions

1. **Weeks** Standard FEM solver, straightforward Metropolis-Hastings dynamics using Langevin move.

Time to converge (in expectation) in EIT problem

Using calibrated forward map, complete electrode model, Gaussian smoothness prior in 2-dimensions

1. **Weeks** Standard FEM solver, straightforward Metropolis-Hastings dynamics using Langevin move.
2. **Days** Fast Jacobian calculation, fast stiffness-matrix assembly exploiting rank-2 local stiffness matrix.

Time to converge (in expectation) in EIT problem

Using calibrated forward map, complete electrode model, Gaussian smoothness prior in 2-dimensions

1. **Weeks** Standard FEM solver, straightforward Metropolis-Hastings dynamics using Langevin move.
2. **Days** Fast Jacobian calculation, fast stiffness-matrix assembly exploiting rank-2 local stiffness matrix.
3. **Hours** Two-step MCMC using linear approximation to forward map

Time to converge (in expectation) in EIT problem

Using calibrated forward map, complete electrode model, Gaussian smoothness prior in 2-dimensions

1. **Weeks** Standard FEM solver, straightforward Metropolis-Hastings dynamics using Langevin move.
2. **Days** Fast Jacobian calculation, fast stiffness-matrix assembly exploiting rank-2 local stiffness matrix.
3. **Hours** Two-step MCMC using linear approximation to forward map
4. $O(1)$ **calculation: Minutes ?** Approximate calculation of forward map, only, with separation time greater than run length

Time to converge (in expectation) in EIT problem

Using calibrated forward map, complete electrode model, Gaussian smoothness prior in 2-dimensions

1. **Weeks** Standard FEM solver, straightforward Metropolis-Hastings dynamics using Langevin move.
2. **Days** Fast Jacobian calculation, fast stiffness-matrix assembly exploiting rank-2 local stiffness matrix.
3. **Hours** Two-step MCMC using linear approximation to forward map
4. $O(1)$ **calculation: Minutes ?** Approximate calculation of forward map, only, with separation time greater than run length
5. **Near real time ?** More computers running something other than MatLab

A Conjugate Direction Sampler for Gaussian Random Fields with Generalization

Conjugate Gradient Optimization

- Basic CG algorithm was introduced by Hestenes and Stiefel (1952)
- Solves the matrix equation $Ax = b$ where A is a symmetric positive definite (SPD) matrix.
- Equivalently minimizes the quadratic form

  $$Q\left(x\right) = \frac{1}{2} x^{\mathsf{T}} A x - b^{\mathsf{T}} x$$

  using a sequence of search directions derived from the gradient of Q projected onto the space conjugate to all previous directions.
- Terminates in $< \dim(x)$ steps (exact arithmetic)
- Actually in $<$ number distinct eigenvalues of A (Jennings 1977)
- Requires storage of two vectors, and one operation by A per iteration.

Efficient sampling from a Gaussian

To sample from a Gaussian distribution with density

$$\pi\left(x\right)=\sqrt{\frac{\det\left(A\right)}{2\pi^{m}}}\exp\left\{-\frac{1}{2}x^{\mathsf{T}}Ax\right\}$$

where $A\in R^{m\times m}$ is SPD, $\dim\left(x\right)=m$. Denote $A^{-1}=\Sigma$.

Standard efficient samplers use the Cholesky factorization

- of covariance Σ (classical algorithm)

  $$RR^{\mathsf{T}}=\Sigma$$

  then if $z\sim N\left(0,I_{m}\right)$, $x=Rz\sim N\left(0,\Sigma\right)$

- of precision matrix A (Rue 2000)

  $$LL^{\mathsf{T}}=A$$

  then if $z\sim N\left(0,I_{m}\right)$, solving $z=L^{\mathsf{T}}x$ gives $x\sim N\left(0,A^{-1}\right)$.

  Markov RF $\Rightarrow$ A is sparse, allowing computationally efficient algorithms.

Generally, if $z\sim N\left(\mu,\Sigma\right)$ then $x=Sz\sim N\left(S\mu,S\Sigma S^{\mathsf{T}}\right)$.

Mutually conjugate vectors (and factorizations)

Definition 1 *A set of non-zero vectors $s^{(i)}$, $i = 0, 1, \ldots, m-1$, is mutually conjugate w.r.t. A (SPD) if*

$$s^{(i)\mathsf{T}} A s^{(j)} = 0 \quad \forall i \neq j.$$

Note that $s^{(i)\mathsf{T}} A s^{(i)} > 0 \ \forall i$, as A is strictly positive definite. Denote $s^{(i)\mathsf{T}} A s^{(i)} = d_i$, hence, if $S \in R^{m \times m}$ is the matrix with columns $s^{(i)}$, $i = 0, 1, \ldots, m-1$, and $D = diag\left(d_0, d_1, \ldots, d_{m-1}\right)$ (also pd), then

$$S^{\mathsf{T}} A S = D.$$

Hence if $z \sim N\left(0, I_m\right)$, $y = \sqrt{D^{-1}} z \sim N\left(0, D^{-1}\right)$, and $x = Sy \sim N\left(0, \Sigma\right)$. This follows since $S^{\mathsf{T}} A S = D$ so $A^{-1} = S D^{-1} S^{\mathsf{T}}$.

The expression for x can be written $x = \sum_{i=0}^{m-1} \left(z_i / \sqrt{d_i}\right) s^{(j)}$. Since the z_i are i.i.d. $\sim N\left(0, 1\right)$, this shows how a sample from π may be generated using a sequence of standard normal random numbers.

Some examples you know of

Example 2 *Cholesky factorizaton of Σ. Since $RR^{\mathsf{T}} = \Sigma$, $R^{\mathsf{T}}AR = I$, so the columns of R are mutually conjugate w.r.t. A.*

Example 3 *Cholesky factorizaton of A. Since $LL^{\mathsf{T}} = A$, $\left(L^{\mathsf{T}}\right)^{-1} L^{-1} = A^{-1}$, so $\left(L^{\mathsf{T}}\right)^{-1} = R$ and so the columns of $\left(L^{\mathsf{T}}\right)^{-1}$ are mutually conjugate w.r.t. A.*

Example 4 *Eigen decomposition of A or Σ. Consider*

$$A = UDU^{\mathsf{T}}$$

where U is unitary (columns are normalised eigenvectors) and D is diagonal (eigenvalues on diagonal). Then $U^{\mathsf{T}}AU = D$, so the eigenvectors are mutually conjugate w.r.t. A.

These examples show that the algorithms based on the Cholesky and eigen factorizations are examples of the more general notion of mutually conjugate w.r.t. A, and also establishes that these algorithms do indeed sample from π.

Some properties

Lemma 5 *The elements of any set of mutually conjugate vectors are linearly independent.*

Proof. Suppose that $\sum \alpha_i s^{(i)} = 0$. Then $0 = \left(\sum \alpha_i s^{(i)}\right)^{\mathsf{T}} A s^{(j)} = \alpha_j d_j \Longrightarrow \alpha_j = 0$. ■

Proposition 6 *Let $\left\{s^{(i)}\right\}_{i=0,1,\ldots,n-1}$ be a set of n mutually conjugate vectors w.r.t. A. A sample from the conditional distribution*

$$\pi\left(x | x \in \textsf{span}\left(s^{(0)}, s^{(1)}, \ldots, s^{(n-1)}\right)\right)$$

may be generated by a sequence of samples from the one-dimensional conditional distributions in directions $s^{(0)}, s^{(1)}, \ldots, s^{(n-1)}$.

Thus, a sample from $\pi\left(x\right)$ may be generated by a sequence of exactly m independent univariable samples taken from the conditional distributions in directions $s^{(0)}, s^{(1)}, \ldots, s^{(m-1)}$ with components in previous directions fixed.

Proof. First consider a sample from the whole space, i.e. $\text{span}\left(s^{(0)}, s^{(1)}, \ldots, s^{(m-1)}\right)$,

$$x = \sum_{i=0}^{m-1} \alpha_i s^{(i)}.$$

From above, $x \sim N\left(0, \Sigma\right)$ when $\alpha \sim N\left(0, D^{-1}\right)$ hence $\alpha_k \sim N\left(0, 1/d_k\right)$ are independent Gaussian. Now consider the conditional distribution along line $s^{(k)}$ when $\alpha_0, \alpha_1, \ldots, \alpha_k$ are fixed, i.e. for

$$\pi\left(x | x \in \alpha_0 s^{(0)} + \alpha_1 s^{(1)} + \cdots \alpha_{k-1} s^{(k-1)} + \text{span}\left(s^{(k)}\right)\right).$$

Write $x = \alpha_0 s^{(0)} + \alpha_1 s^{(1)} + \cdots \alpha_{k-1} s^{(k-1)} + \lambda s^{(k)}$. How is λ distributed?

$$\begin{aligned} \pi\left(\lambda\right) &\propto \exp\left\{-\frac{1}{2}\left(\sum_{i=0}^{k-1} \alpha_i s^{(i)} + \lambda s^{(k)}\right)^{\mathsf{T}} A \left(\sum_{i=0}^{k-1} \alpha_i s^{(i)} + \lambda s^{(k)}\right)\right\} \\ &= \exp\left\{-\frac{1}{2}\left(\sum_{i=0}^{k-1} \alpha_i^2 d_i + \lambda^2 d_k\right)\right\} \end{aligned}$$

i.e. $\lambda \sim N\left(0, 1/d_k\right)$ which is the same as the distribution of α_k in a sample from $\pi\left(x\right)$. ■

Auxiliary vector and Quadratic Form

Given $x = \alpha_0 s^{(0)} + \alpha_1 s^{(1)} + \cdots \alpha_{k-1} s^{(k-1)}$ define $b^{(k)} = \alpha_0 A s^{(0)} + \alpha_1 A s^{(1)} + \cdots \alpha_{k-1} A s^{(k-1)}$.

Proposition 7 $x^{(k)} = \alpha_0 s^{(0)} + \alpha_1 s^{(1)} + \cdots \alpha_{k-1} s^{(k-1)}$ *minimises the quadratic form*

$$Q_k\left(x\right) = \frac{1}{2} x^{\mathsf{T}} A x - b^{(k)\mathsf{T}} x$$

in the subspace span $\left(s^{(0)}, s^{(1)}, \ldots, s^{(k-1)}\right)$.

Proof. The minimum occurs when for $j = 0, 1, \ldots, k-1$

$$\begin{aligned} 0 &= \frac{\partial Q_k\left(\sum_{i=0}^{k-1} a_i s^{(i)}\right)}{a_j} \\ &= \frac{\partial}{a_j}\left(\frac{1}{2}\left(\sum_{i=0}^{k-1} a_i s^{(i)}\right)^{\mathsf{T}} A \left(\sum_{i=0}^{k-1} a_i s^{(i)}\right) - \left(\sum_{i=0}^{k-1} \alpha_i A s^{(i)}\right)^{\mathsf{T}} \left(\sum_{i=0}^{k-1} a_i s^{(i)}\right)\right) \\ &= \frac{\partial}{a_j}\left(\frac{1}{2}\sum_{i=0}^{k-1} a_i^2 d_i - \sum_{i=0}^{k-1} \alpha_i a_i d_i\right) = a_j d_j - \alpha_j d_j. \end{aligned}$$

■

Generating mutually conjugate vectors

Given any set of linearly independent independent vectors $\left\{r^{(i)}\right\}_{k=0,2,\ldots,m-1}$, we can form a mutually conjugate set $\left\{s^{(i)}\right\}_{k=0,2,\ldots,m-1}$ (spanning the same space) using a modified Gram-Schmidt process. First set $s^{(0)} = r^{(0)}$ and then define

$$s^{(k)} = r^{(k)} - \sum_{i=0}^{k-1} \frac{r^{(k)\mathsf{T}} A s^{(i)}}{d_i} \quad \text{for } k = 1, 2, \ldots, m-1. \tag{1}$$

It is easy to check that the set $\left\{s^{(i)}\right\}_{k=0,2,\ldots,m-1}$ has the right properties.

This requires storage of all previous vectors to evaluate the summation. However, if the set $\left\{r^{(i)}\right\}_{k=0,2,\ldots,m-1}$ is chosen carefully, most of the terms in the summation evaluate to zero, e.g.

$$\begin{aligned} r^{(k)} &= -\nabla Q_k\left(x^{(k)}\right) = b^{(k)} - Ax^{(k)} \\ &= \nabla \log \pi\left(x^{(k)}\right) + b^{(k)}. \end{aligned} \tag{2}$$

Naive algorithm

Algorithm 8 *Set initial values for $x^{(0)}$ and $b^{(0)}$ (not both zero)*

Initialize $s^{(0)} = \nabla \log \pi \left(x^{(0)}\right) + b^{(0)} = r^{(0)} = b^{(0)} - Ax^{(0)}$*. Drawing a sample from the conditional distribution*

$$\pi \left(x|x \in x^{(0)} + span \left(s^{(0)}\right)\right)$$

gives the next point $x^{(1)} = x^{(0)} + \left(\alpha_0 - \frac{s^{(0)\mathsf{T}}Ax^{(0)}}{d_0}\right) s^{(0)}$, $b^{(1)} = b^{(0)} + \left(\alpha_0 - \frac{s^{(0)\mathsf{T}}Ab^{(0)}}{d_0}\right) As^{(0)}$

Given the current point $x^{(k)}$*, evaluate new residual vector* $r^{(k)}$ *is using 2. The sampling direction* $s^{(k)}$ *is calculated and 1 so as to be conjugate to all previous sampling directions. Conditional sampling from* $x^{(k)}$ *in direction* $s^{(k)}$ *may be achieved by drawing* $\alpha_k \sim N\left(0, 1/d_k\right)$ *and setting*

$$x^{(k+1)} = x^{(k)} + \left(\alpha_k - \frac{s^{(k)\mathsf{T}}Ax^{(k)}}{d_k}\right) s^{(k)}$$

the auxiliary vector is updated as

$$b^{(k+1)} = b^{(k)} + \left(\alpha_k - \frac{s^{(k)\mathsf{T}}b^{(k)}}{d_k}\right) As^{(k)}.$$

Some nifty results

Lemma 9 *The vector* $r^{(k)}$ *is orthogonal to the space span* $\left(s^{(0)}, s^{(1)}, \ldots, s^{(k-1)}\right)$*.*

Proof. By proposition 2, $x^{(k)}$ is the point in the space span $\left(s^{(0)}, s^{(1)}, \ldots, s^{(k-1)}\right)$ at which Q_k is minimised. Hence the projection of the gradient $r^{(k)} = -\nabla Q_k\left(x^{(k)}\right)$ onto the space is zero, i.e. $r^{(k)}$ is orthogonal to the space. ■

Lemma 10 ${r^{(i)}}^T r^{(j)} = 0$ *for all* $i \neq j$*.*

Proof. By construction, span $\left(s^{(0)}, s^{(1)}, \ldots, s^{(k-1)}\right)$ is the same space as span $\big(r^{(0)}, r^{(1)}, \ldots, r^{(k-}$ hence from last Lemma, $r^{(k)}$ is orthogonal to $r^{(j)}$ for all $j < k$. The result follows from symmetry. ■

The key nifty result

Proposition 11 *In algorithm 1, all the terms in the summation in 1 vanish, except for* $i = k - 1$.

Proof. Consider the case $x^{(0)} = 0$, so the update rule is $x^{(k+1)} = x^{(k)} + \alpha_k s^{(k)}$. Hence

$$\begin{aligned} r^{(i+1)} &= b^{(i+1)} - Ax^{(i+1)} \\ &= b^{(i)} + \left(\alpha_i - \frac{s^{(i)\mathsf{T}} b^{(i)}}{d_i}\right) As^{(i)} - A\left(x^{(i)} + \alpha_i s^{(i)}\right) \\ &= r^{(i)} + \frac{s^{(i)\mathsf{T}} b^{(i)}}{d_i} As^{(i)}. \end{aligned}$$

So $As^{(i)} = \left(r^{(i+1)} - r^{(i)}\right) \frac{d_i}{s^{(i)\mathsf{T}} Ab^{(i)}}$. Hence the summation is

$$\sum_{i=0}^{k-1} \frac{r^{(k)\mathsf{T}} As^{(i)}}{d_i} = \sum_{i=0}^{k-1} \frac{r^{(k)\mathsf{T}} \left(r^{(i+1)} - r^{(i)}\right)}{s^{(i)\mathsf{T}} b^{(i)}} = \frac{r^{(k)\mathsf{T}} r^{(k)}}{s^{(k-1)\mathsf{T}} b^{(k-1)}}.$$

Note: including full update for $x^{(i+1)}$ leaves $As^{(i)}$ a scalar multiple of $r^{(i+1)} - r^{(i)}$, hence proposition holds with modified expression for sum. ■

The algorithm

Algorithm 12 *initialise* $x^{(0)}$ *and* $b^{(0)}$ *(not both zero)*

```
$r^{(0)} = b^{(0)} - Ax^{(0)}$
$s^{(0)} = r^{(0)}$
for $k = 1$ to $m$ do
begin
    $d_{k-1} = s^{(k-1)\mathsf{T}} A s^{(k-1)}$
    $\alpha_{k-1}\tilde{}N\left(0,1\right)/\sqrt{d_{k-1}}$
    $x^{(k)} = x^{(k-1)} + \left(\alpha_{k-1} - \frac{s^{(k-1)\mathsf{T}} A x^{(k-1)}}{d_{k-1}}\right) s^{(k-1)}$
    $b^{(k)} = b^{(k-1)} + \left(\alpha_{k-1} - \frac{s^{(k-1)\mathsf{T}} b^{(k-1)}}{d_{k-1}}\right) A s^{(k-1)}$
    $r^{(k)} = b^{(k-1)} - Ax^{(k-1)}$
    $s^{(k)} = r^{(k)} - \frac{r^{(k)\mathsf{T}} r^{(k)}}{s^{(k-1)\mathsf{T}} b^{(k-1)}}$
end
```

A bonus!

Algorithm generates $x^{(n)}\tilde{}N\left(0, A^{-1}\right)$

Lemma 13 *If* $\left\{s^{(i)}\right\}_{k=0,2,\ldots,m-1}$ *are mutually conjugate w.r.t.* A *(spd), then* $\left\{q^{(i)}\right\}_{k=0,2,\ldots,m-1} =$ $\left\{As^{(i)}\right\}_{k=0,2,\ldots,m-1}$ *are mutually conjugate w.r.t.* A^{-1}*.*

Proof. $Q^{\mathsf{T}}A^{-1}Q = (AS)^{\mathsf{T}}A^{-1}(AS) = S^{\mathsf{T}}A^{\mathsf{T}}A^{-1}AS = S^{\mathsf{T}}AS = D.$ ■

It follows that $b^{(n)}\tilde{}N(0, A)$, i.e. with A as covariance matrix rather than precision matrix.

Generalization for non-Gaussian distributions

To sample from $\pi(\cdot)$, $f = \log \pi$

```
Algorithm 14  initialize $x_0$, $b_0$ ($b_0 \neq \nabla f(x_0)$)
Evaluate $r_0 = b_0 - \nabla f(x_0)$
Set $p_0 = r_0$
$k = 0$
while $\|r_k\| \neq 0$ do
    sample conditional density $x_{k+1} \sim \pi(x_{k+1}\textit{span}\{p_k\})$
    Evaluate $\nabla f(x_{k+1})$
    Set $b_{k+1} = b_k + \gamma_k p_k$
    Set $r_{k+1} = b_{k+1} - \nabla f(x_{k+1})$
    Set $p_{k+1} = r_{k+1} + \beta_{k+1} p_k$
    $k = k + 1$
end (while)
```

Conclusions

1. There is much work still to make Bayesian inference a robust off-the-shelf solution to inverse problems
2. Time to convergence depends on efficient simulation of the forward map, fast MCMC algorithms, good proposal distributions
3. BMIP fit into the Physics intuition easily (and much was developed by Physicists)
4. Deterministic sampling algorithms?