ELEC 446 Computational Inference
Assignment 2

Due Monday 5 September (after mid-semester break) by close of play. This assignment is worth 10% of your final grade.

Your task is to apply the MTC algorithm (Option 1) in Fox and Norton 2016 (copy on 446 website) to perform deblurring by sample-based Bayesian inference, and hence sharpen up a photograph of Jupiter – the same one you deblurred using regularization in ELEC445 (if you did that).

You will find the following files on the 446 website (`https://coursesupport.physics.otago.ac.nz/wiki/pmwiki.php/ELEC446`): (Please right click these to 'Save As', to get formatting correct.)

1. `jupiter1.tif` which contains a photograph of Jupiter taken in the methane band (780nm) on a grid of size 256×256 pixels, each takes an integer value from 0 to 255. In the upper right-hand portion of the photograph is one of the Galilean satellites. This satellite is effectively a point source, and so we can use that region of the photograph as the point-spread function k.
2. `jupiter.m` contains Matlab code to read in the photograph, display it, and extract a 32×32 window of points around the satellite as the point spread function.
3. `deconv1.m` is code for performing regularized deconvolution.

A Précis of the Model

Denoting the true unknown image by x and the data by y, the observation process is

$$y = k * x + \eta = Ax + \eta$$

where the forward map A is the linear operator representing convolution with the point-spread function k, and η is an unknown 'noise' vector representing measurement errors including digitization. We will assume the additive noise is iid zero mean Gaussian noise with unknown precision (inverse of variance) γ, i.e. $\eta \sim \mathrm{N}\left(0, \gamma^{-1} I\right)$. Hence the distribution over y conditioned on image x and precision γ is

$$y|x, \gamma \sim \mathrm{N}\left(Ax, \gamma^{-1} I\right).$$

We can prefer smoothness in the unknown 'true' image x by modeling x as a draw from a Gaussian Markov random field (GMRF) that assigns low probability to non-smooth images. We use the (intrinsic) multivariate Gaussian

$$x|\delta \sim \mathrm{N}(0, (\delta L)^{-1})$$

as prior distribution (stochastic image model) where the matrix L is the discrete Laplacian, equivalent to convolution with the 5-point finite-difference stencil, i.e.,

$$Lx \equiv x * \begin{pmatrix} & -1 & \\ -1 & 4 & -1 \\ & -1 & \end{pmatrix}.$$

We will assume that each of x and y are periodic (in each dimension) so that convolution is the same as circular convolution, i.e., $Ax \equiv k \circledast x$, $Lx \equiv s \circledast x$ (s is the 5-point stencil). Then the operation of A and L may be computed by Fourier multiplication, without zero padding. Not zero padding x and y is important for efficiency in the sampling method you will use, because then A and L are diagonalized by the (discrete) Fourier transform and so the eigenvalues of $A^T A$ and L are directly available. Denote (discrete) Fourier transform pairs by $\leftrightarrow$, evaluated by the FFT, so $k \leftrightarrow K$, $s \leftrightarrow S$, etc. Note that in image space, k, s, etc, are indexed by two-dimensional indices, say i and j that are horizontal and vertical positions of pixels in the images, and range $i, j = 1, 2, \ldots, 256$ after zero-padding to image size. The Fourier transforms are also indexed by two-dimensional indices, say $l, m = 1, 2, \ldots, 256$, with units of spatial frequency (inverse length). In the Fourier domain the forward map and prior operators are

$$\begin{aligned} Ax = k \circledast x &\quad \leftrightarrow K \times X \\ Lx = s \circledast x &\quad \leftrightarrow S \times X \end{aligned}$$

In these formulas the multiplication is *component-wise* (`.*` in matlab) – images are arrays; matrix operations make no sense.

The matrix L is symmetric ($L^T = L$) when the stencil s is suitably centred, i.e. with the 4 at the origin, which is the $(1, 1)$ position in matlab. Then the eigenvalues of L are real, and are exactly the components of S (that is real and even because s is real and even; symmetry property §3.3.10 in Linear System notes). When the stencil is not centred, the components of S change by a phase (shifting property §3.3.4 in Linear System notes) and the eigenvalues of L are the magnitude $|S|$. Similarly, the eigenvalues of $A^T A$ are given by $|K|^2$ (more on this later). In particular, determinants may be computed as the product of eigenvalues. Since L and $A^T A$ are simultaneously diagonalized by the Fourier transform (the two operators have the same eigenvectors), the crucial determinant $\det(\lambda L + A^T A)$ is simply the product of components in $(\lambda|S| + |K|^2)$. Note also that the argument of exponentials may be computed efficiently in the Fourier domain by utilizing Parseval's theorem.

The Bayesian model is succinctly written in the hierarchical form

$$\begin{aligned} y|x, \gamma &\sim \mathrm{N}\left(Ax, \gamma^{-1} I\right) \\ x|\delta &\sim \mathrm{N}\left(0, (\delta L)^{-1}\right) \\ (\gamma, \delta) &\sim \pi\left(\gamma, \delta\right) \end{aligned}$$

where $\pi(\gamma, \delta)$ denotes the (hyper)prior over the unknown parameters γ and δ. This gives the posterior distribution over unknowns x and γ, δ conditioned on measured y, given by Bayes' rule as

$$\pi(x, \gamma, \delta|y) \propto \pi(y|x, \gamma, \delta)\, \pi(x, \gamma, \delta) = \pi(y|x, \gamma, \delta)\, \pi(x|\gamma, \delta)\, \pi(\gamma, \delta)\,.$$

which is just the product of the three probability density functions in the hierarchical model.

For (hyper)prior distribution over γ and δ you may use the Gamma distributions defined in Fox and Norton Eqns 17 and 18, or simply use the (improper) distribution that restricts γ and δ to be positive, whichever is simpler for you.

You will implement the MTC sampler that first draws *independent* (γ, δ) from the marginal posterior for (γ, δ) given y,

$$(\gamma, \delta) \overset{\text{iid}}{\sim} \pi(\gamma, \delta|y)$$

then draws from the full conditional for x (using the γ, δ sample from the previous step)

$$x \sim \pi(x|y, \gamma, \delta)$$

to generate an *independent* sample from the posterior distribution. Many independent (γ, δ) allows you to generate many independent posterior samples (x, γ, δ) (given y).

Assignment Steps

1. (Some theory work around discrete Fourier transforms)

   (a) Given that $Ax = k \circledast x$, show that $A^T y = h \circledast y$ where $h(t) = k(-t)$ (t is a two-dimensional index).

   (b) Hence show that $A^T y \leftrightarrow K^* \times Y$ (* denotes componentwise conjugate).

   (c) For an $N \times N$ array $z(l,m)$, the 2-dimensional discrete Fourier transform is

   $$Z[p,q] = \frac{1}{N^2} \sum_{l=0}^{N-1} \sum_{m=0}^{N-1} z(l,m) \exp\left(-j2\pi \frac{lp+mq}{N}\right)$$

   and the inverse transform is

   $$z(l,m) = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} Z[p,q] \exp\left(j2\pi \frac{lp+mq}{N}\right).$$

   Note that both z and Z have period N in each index. Note that, if s is real and even (hence L is symmetric), S is also real and even. Show that the eigenvalues of L are $S[p,q]$ and eigenvalues of $A^T A$ are $|K|^2$. What are the corresponding (real) eigenvectors?

2. (Sampling from the marginal posterior over hyperparameters) The marginal posterior for (γ, δ) given y, is

   $$\pi(\gamma, \delta | y) \propto \sqrt{\frac{\det(\gamma I)\det(\delta L)}{\det(\delta L + \gamma A^T A)}} \exp\left\{-\frac{1}{2}\gamma y^T \left[I - \gamma A(\delta L + \gamma A^T A)^{-1} A^T\right] y\right\} \pi(\theta)$$

   up to a normalizing constant not depending on γ or δ.

   (a) Derive this expression for $\pi(\gamma, \delta|y)$, perhaps using identity in Lemma 2 of Fox and Norton 2016. (You might find it helpful to do 3a first!)

   (b) Assuming a symmetric proposal, write down the Metropolis-Hasting acceptance probability used in an MCMC for sampling from this distribution. (Cancel as many determinants as possible.)

   (c) Show that in the Fourier domain, determinants may be calculated as the product of the magnitude of the Fourier transform of the point spread function plus a multiple of the Fourier transform of the '5-point' stencil.

   (d) Write a program to implement adaptive Metropolis MCMC that samples from this distribution. Use Fourier transforms (evaluated once when initializing your program) to simplify the calculation of the required determinants, and argument of the exponential. Generate, say, 10000, samples – as long as that does not take too long.

   (e) Plot histograms of the marginal posterior distributions for each of γ, δ and the effective regularizing parameter $\lambda = \delta/\gamma$.

   (f) Give the integrated autocorrelation time for the parameter λ. How many 'effectively independent' samples have you generated?

3. (Sampling from the full conditional for x) The full conditional for x is

   $$x|\gamma, \delta, y \sim \mathrm{N}\left((\gamma A^T A + \delta L)^{-1} \gamma A^T y, (\gamma A^T A + \delta L)^{-1}\right).$$

   An independent sample from this distribution may be computed by solving

   $$\left(\gamma A^T A + \delta L\right) x = \gamma A^T y + w$$

   where $w = v_1 + v_2$ with independent $v_1 \sim \mathrm{N}\left(0, \gamma A^T A\right)$ and $v_2 \sim \mathrm{N}(0, \delta L)$. Give the equivalent expression in terms of Fourier transforms.

(a) Give expressions for the probability density functions for $\pi(y|x,\gamma) = \mathrm{N}\left(Ax, \gamma^{-1}I\right)$ and $\pi(x|\delta) = \mathrm{N}\left(0, (\delta L)^{-1}\right)$, and hence, by Bayes' rule, derive the expression for the pdf of the full conditional $\pi(x|\gamma,\delta,y)$.

(b) Write program segments to sample each of $v_1 \sim \mathrm{N}\left(0, \gamma A^T A\right)$ and $v_2 \sim \mathrm{N}\left(0, \delta L\right)$ by suitably transforming standard normal random variables. The action of A may be computed in the Fourier domain, to give the Fourier transform of v_1. You will need a square-root of L, which is easy because L is real-symmetric and so the Fourier transform is real, as long as you centered the 5-point stencil correctly.

(c) Solve the equation above, to generate a sample from the full conditional for x. Note that the equation to be solved is very close to a randomized version of the generalized deconvolution equations that you solved in Assignment 1, so you can reuse your code that performed the solve in the Fourier domain. (Note that we are using Laplacian regularization here, rather than Tikhonov as in Assignment 1.)

(d) Plot a (independent) sample from the posterior distribution over deblurred images.

(e) Use 100 independent posterior samples to compute and plot a mean image.

(f) Is your posterior mean image better, worse, the same as, the regularized deconvolved image you produced in Assignment 1?

Challenge question (for interest only): You will see in your deblurred image that there are some non-physical ripples around the deconvolved Galilean satellite. Can you improve the reconstruction by using a simple parametric model for the point-spread function and inferring and averaging over those parameters?