DOE/NV/25946--2491

A Metropolis-within-Gibbs sampler for MCMC uncertainty quantifications in PSF reconstruction

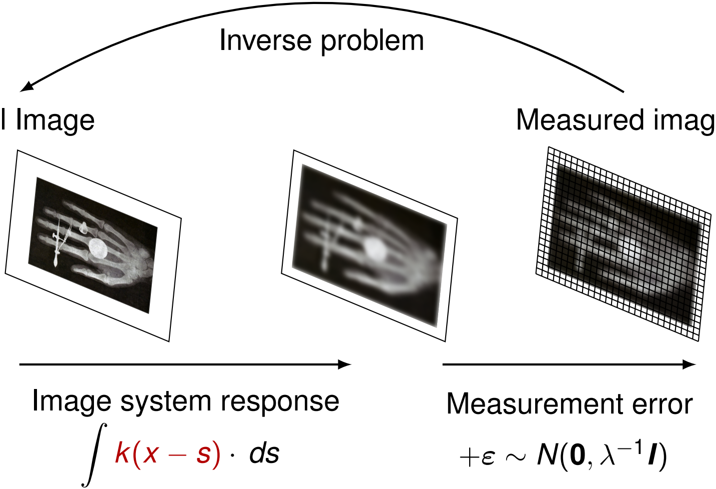
John Bardsley[†], Kevin Joyce[†*], Aaron Luttman[*]

[†]The University of Montana
[*]National Security Technologies LLC

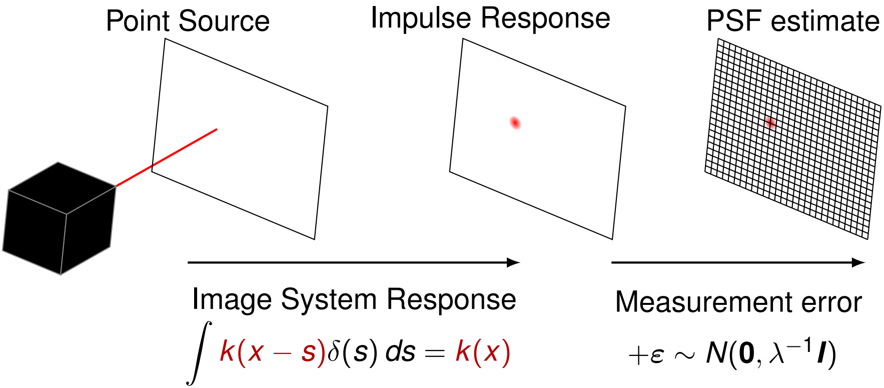
MUQ[**] June 24, 2015

This work was done by National Security Technologies, LLC, under Contract No. DE-AC52-06NA25946 with the U.S. Department of Energy and supported by the Site-Directed Research and Development Program.

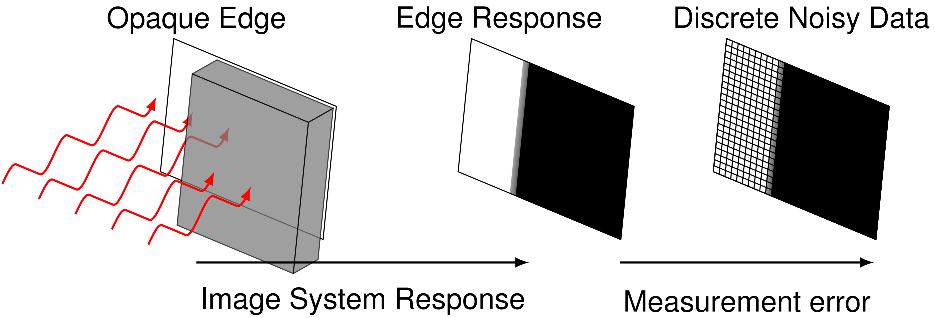
Stochastic Imaging Model



Point Spread Function Estimation



Point Spread Function Estimation

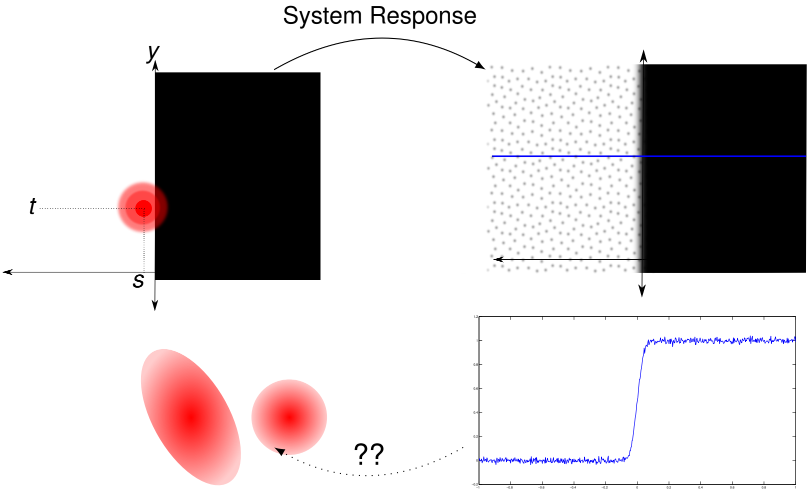


$$f(x) = \int k(x-s) E(s)\, ds$$

$$+\varepsilon \sim N(\mathbf{0}, \lambda^{-1} \boldsymbol{I})$$

Edge blur problem

$$f(x_i) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(s,t)E(x_i - s)dtds + \varepsilon_i$$



Radial Symmetry

- Assume $k(x, y) = \rho(\sqrt{x^2 + y^2}) = \rho \circ T(x, y)$
- Precomposition can be thought of as an operator $k = T^\sharp \rho$
- We assume k is in the range of $T^\sharp$, then reformulate the inverse problem on ρ.

$$f = Bk \quad \text{if and only if} \quad f = BT^\sharp \rho$$

- And prior assumptions about k are directly translated to ρ through $T^\sharp$. I.e.

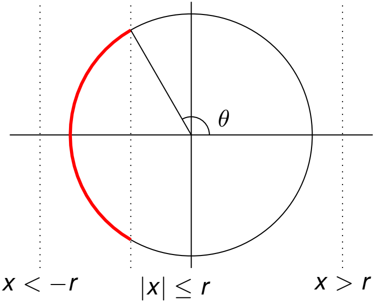
$$k \sim N(0, \delta^{-1} Q) \quad \text{if and only if} \quad T^\sharp k \sim N(0, \delta^{-1} Q)$$

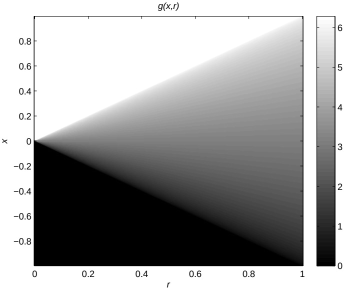
- The work on this theory is ongoing.

Radially Symmetric PSF

If we assume radial symmetry, $k(x,y) = \rho(\sqrt{x^2+y^2})$, and that the edge is indicated at $x = 0$ by E, then

$$\begin{aligned} f(x_i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(s,t)E(x_i - s)dtds + \varepsilon_i \\ &= \int_0^{\infty} \rho(r)\left(\int_0^{2\pi} E(x_i - r\cos\theta)d\theta\right) rdr + \varepsilon_i \\ &= \int_0^{\infty} \rho(r)\cdot g(x_i,r)\, rdr + \varepsilon_i. \end{aligned}$$





Observe that g is symmetric about the $x = 0$.

Bayesian Stochastic Inverse Problem

For the inverse problem

$$\boldsymbol{f} = \mathcal{G}\rho + \boldsymbol{\epsilon},$$

regularization and uncertainty quantification can be achieved by modeling the PSF as a random quantity with an appropriate prior.

- Measurement error is given by Gaussian likelihood with $\mathrm{E}\,\boldsymbol{f} = \mathcal{G}\rho$ and $\mathrm{Var}\,\boldsymbol{f} = \lambda^{-1}\boldsymbol{I}$.
- We use a spatial automodel Gaussian prior defined implicitly on k. Denote Δ_0 as the Laplacian with zero boundary conditions, then if k is such that $\mathrm{E}\,\Delta_0^m k = 0$ and $\mathrm{Var}\,\Delta_0^m k = \delta^{-1} I$, this enforces a prior assumption of “smoothness” for k.
- This prior translates to the radial representation as $\delta_0^m k = \Delta_0^m T^\sharp \rho = \Delta_0^m(\rho \circ T)(r) = \left(\frac{1}{r}\frac{d}{dr} r \frac{d}{dr}\right)^2 \rho(r).$
- Note that the symmetry implicitely enforces a left reflective boundary condition.

Bayesian Stochastic Inverse Problem

- The variable λ determines the variance of the measurement error and the variable δ determines the "strength" of the smoothness operator.
- They are assumed to have an "uniformative" Gamma distribution with hyper-parameters $\alpha = 1$ and $\beta = 10^6$ [CS08].

The joint posterior density for (ρ, λ, δ) is given by Bayes' Theorem

$$\begin{aligned} \pi(\rho, \lambda, \delta | \boldsymbol{f}) &= \frac{\pi(\boldsymbol{f} | \rho, \lambda, \delta)\pi(\rho, \lambda, \delta)}{\pi(\boldsymbol{f})} \\ &= \frac{\pi(\boldsymbol{f} | \rho, \lambda)\pi(\rho | \delta, \lambda)\pi(\lambda, \delta)}{\pi(\boldsymbol{f})} \\ &\propto \pi(\boldsymbol{f} | \rho, \lambda)\pi(\rho | \delta)\pi(\lambda, \delta) \end{aligned}$$

Hierarchical Gibbs Sampler

- Let $\boldsymbol{\rho}$ be discretized with n points and $\boldsymbol{f}$ with $2n$ points.
- Denote the discretization of the forward operator $BT^\sharp$ with $\boldsymbol{G}$ and the discretization of the radial Laplacian with $\boldsymbol{L}$.
- Then $\pi(\boldsymbol{f}|\boldsymbol{\rho},\lambda) \propto \lambda^n \exp\left(-\frac{\lambda}{2}\|\boldsymbol{f}-\boldsymbol{G}\boldsymbol{\rho}\|^2\right)$,
- and $\pi(\boldsymbol{\rho}|\delta) \propto \delta^{n/2} \exp\left(-\frac{\delta}{2}\langle\boldsymbol{\rho},\boldsymbol{L}^m\boldsymbol{\rho}\rangle\right)$,
- and $\pi(\lambda,\delta) \propto (\lambda\delta)^{\alpha-1}\exp(-\beta(\lambda+\delta))$.

It can be shown, by completing the square, that

$$-\frac{\lambda}{2}\|\boldsymbol{f}-\boldsymbol{G}\boldsymbol{\rho}\|^2 - \frac{\delta}{2}\langle\boldsymbol{\rho},\boldsymbol{L}^m\boldsymbol{\rho}\rangle = -\frac{1}{2}\Big\langle \boldsymbol{R}(\boldsymbol{\rho}-\boldsymbol{\mu}),(\boldsymbol{\rho}-\boldsymbol{\mu})\Big\rangle,$$

where $\boldsymbol{R} = (\lambda\boldsymbol{G}^*\boldsymbol{G} + \delta\boldsymbol{L}^{m/2})\boldsymbol{\rho}$ and $\boldsymbol{\mu} = \lambda\boldsymbol{R}^{-1}\boldsymbol{G}^*\boldsymbol{f}$.

- So, $\pi(\boldsymbol{f}|\boldsymbol{\rho},\lambda)\pi(\boldsymbol{\rho}|\delta)$ is the kernel to a multivariate normal with covariance $\boldsymbol{R}$ and mean $\boldsymbol{\mu}$.

Hierarchical Gibbs Sampler

In order to draw efficiently from the conditional density for $\boldsymbol{\rho}$, we use the Cholesky decomposition of $\boldsymbol{R} = \boldsymbol{S}^* \boldsymbol{S}$ which dominates the computation at $O(n^3)$.

The Hierarchical Gibbs sample is:

1: Let δ_0 and λ_0 be given and set $k = 0$.
2: Compute $\boldsymbol{R_k} = \boldsymbol{S_k}^T \boldsymbol{S_k}$.
3: Draw $\boldsymbol{\rho}_k$ from $N\left(\lambda_k \boldsymbol{R_k}^{-1} \boldsymbol{G}^T \boldsymbol{f}, \boldsymbol{R_k}^{-1}\right)$.
4: Draw λ_{k+1} from $\Gamma\left(n + \alpha,\ \frac{1}{2}\|\boldsymbol{G}\boldsymbol{\rho}_k - \boldsymbol{f}\|^2 - \beta\right)$
5: Draw δ_{k+1} from $\Gamma\left(n/2 + \alpha,\ \frac{1}{2}\|\boldsymbol{L}\boldsymbol{\rho}_k\|^2 - \beta\right)$
6: Set $k = k + 1$ and return to 2.

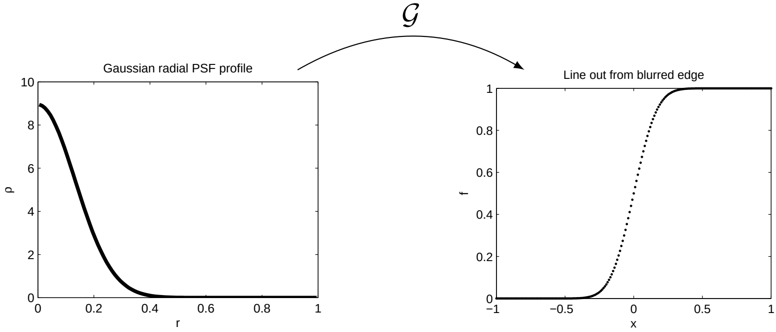
A Synthetic Example

A radially symmetric two-dimensional Gaussian PSF has the form

$$\rho(r_j) = (2\pi\sigma^2)^{-1} e^{\frac{-r_j^2}{2\sigma^2}}$$

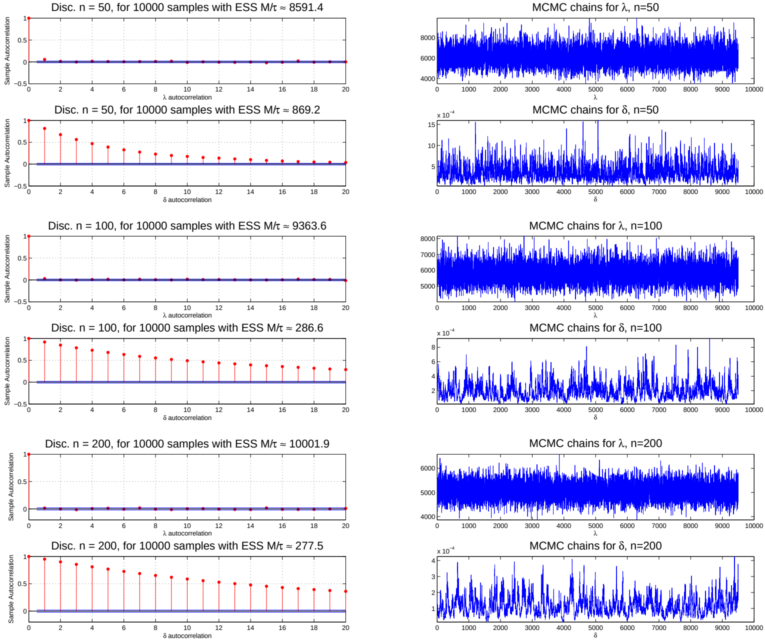
and it can be shown analytically that the analytic forward blur is an error function

$$[\mathcal{G}\rho](x_i) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x_i} e^{-\frac{s^2}{2\sigma^2}}\, ds$$

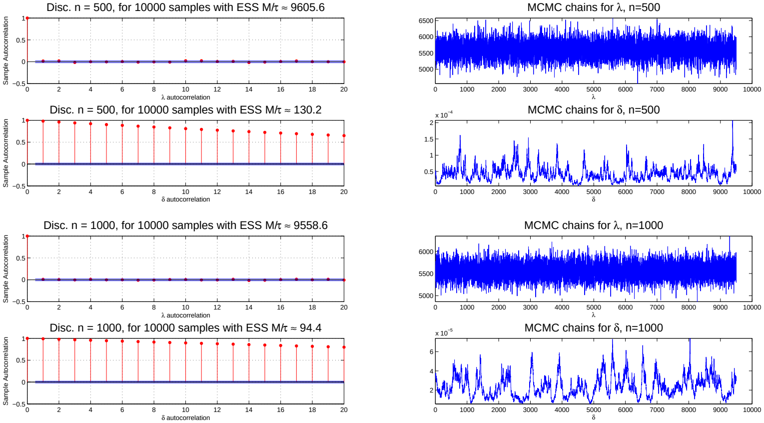


mcmc_hierarchical.png

Correlated δ chains



Correlated δ chains



Marginalized Posterior Density

- As discretizations of $\boldsymbol{\rho}$ become finer, the correlation between the parameter δ and the solution $\boldsymbol{\rho}$ becomes stronger [ABPS14].
- If we marginalize $\pi(\boldsymbol{\rho}, \lambda, \delta | \boldsymbol{f})$ by $\boldsymbol{\rho}$, we obtain $\pi(\lambda, \delta | \boldsymbol{f})$.
- Then $\pi(\lambda, \delta | \boldsymbol{f})\pi(\boldsymbol{\rho} | \lambda, \delta, \boldsymbol{f}) = \pi(\boldsymbol{\rho}, \lambda, \delta | \boldsymbol{f})$.

$$
\begin{aligned}
\int \pi(\boldsymbol{\rho}, \lambda, \delta | \boldsymbol{f}) d\boldsymbol{\rho} &\propto \int (\lambda^2 \delta)^{n/2+\alpha} \exp\left(-\frac{\lambda}{2} \|\boldsymbol{G}\boldsymbol{\rho} - \boldsymbol{f}\|^2 - \frac{\delta}{2} \|\boldsymbol{L}\boldsymbol{\rho}\|^2 - \beta(\lambda - \delta) \right) d\boldsymbol{\rho} \\
&= \int (\lambda^2 \delta)^{n/2+\alpha} \exp\left(-\frac{1}{2} \left\langle \boldsymbol{R}(\boldsymbol{\rho} - \boldsymbol{\mu}), (\boldsymbol{\rho} - \boldsymbol{\mu}) \right\rangle \right. \\
&\qquad \left. -\frac{1}{2} U(\lambda, \delta, \boldsymbol{b}) - \beta(\lambda - \delta) \right) d\boldsymbol{\rho} \\
&= (\lambda^2 \delta)^{n/2+\alpha} c(\lambda, \delta, \boldsymbol{f}) \exp\left(-\frac{1}{2} U(\lambda, \delta, \boldsymbol{b}) - \beta(\lambda - \delta) \right),
\end{aligned}
$$

where $U = \left\langle (\lambda \boldsymbol{I} - \lambda^2 \boldsymbol{G}\boldsymbol{R}^{-1}\boldsymbol{G}^*)\boldsymbol{f}, \boldsymbol{f} \right\rangle$ and $c = \det \boldsymbol{R}^{-1/2}$.

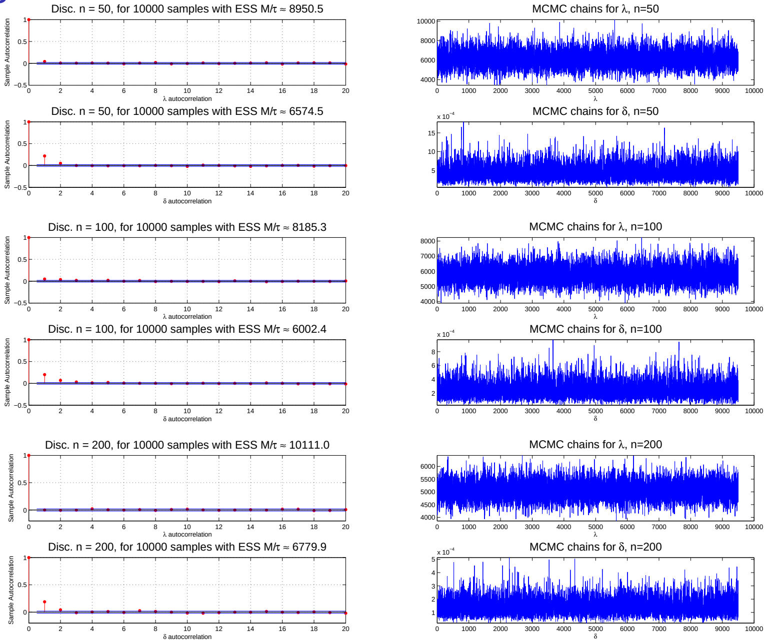
- To draw from this density, we embed a Metropolis-Hastings algorithm within the Gibbs sampler.
- Both U and c can be carried out using a Cholesky factorization ($O(n^3)$), and will be required for each Metropolis step.

Metropolis-Hastings within Gibbs Sampling

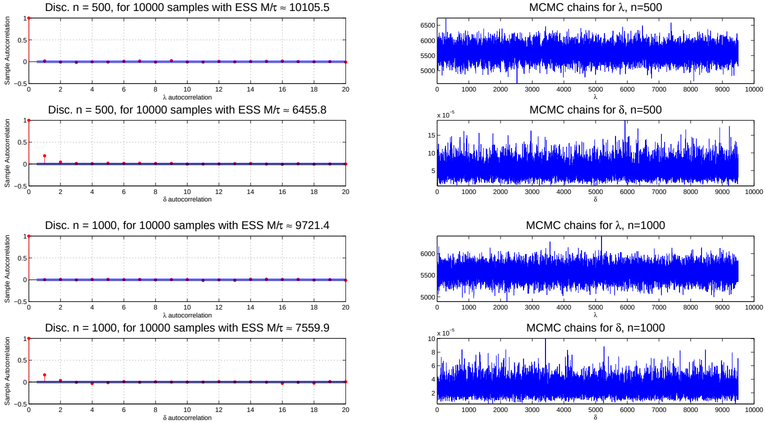
- Since δ is the problematic parameter, we implment the Metropolis-Hastings step on its conditional marginalized density $p(\delta|\boldsymbol{f},\lambda)$.
- We use a Gaussian proposal with adaptive variance. We use an initial variance estimate and support by running a Hierarchical Gibbs sampler.
- Due to numerical overflow issues, all computations are carried out on the log scale.

1: Let λ_0, δ_0 and γ be given.
2: Compute $\boldsymbol{R_k} = \boldsymbol{S_k}^T\boldsymbol{S_k}$.
3: Draw $\boldsymbol{x}_k$ from $N\left(\boldsymbol{R}_k^{-1}\lambda_k\boldsymbol{A}^T\boldsymbol{b}, \boldsymbol{R}_k^{-1}\right)$.
4: Draw λ_{k+1} from $\Gamma\left(n/2+\alpha,\ \frac{1}{2}\|\boldsymbol{Ax_k}-\boldsymbol{b}\|^2-\beta\right)$.
5: Set $j=1$, $\widetilde{\delta}_0 = \frac{1}{k}\sum_{i=1}^k \delta_i$. Compute $c_0 = \log c(\lambda_k, \widetilde{\delta}_0)$ and then $U_0 = \log U(\lambda_k, \widetilde{\delta}_0, \boldsymbol{b})$, and $p_0 = \log p(\widetilde{\delta}_0|\boldsymbol{f}\lambda_k)$.
6: **while** $j < \widetilde{M}$ **do**
7:     Draw $\widetilde{\delta}_j$ from $N(\widetilde{\delta}_{j-1}, \gamma)$.
8:     Compute $U_j = U(\lambda_k, \widetilde{\delta}_j)$, $c_j = c(\lambda_k, \widetilde{\delta}_j)$, and then $p_j = \log p(\widetilde{\delta}_j|\boldsymbol{f}\lambda_k)$.
9:     Draw u_j from $\mathrm{Unif}[0,1]$ and if $p_j - p_{j-1} < \log u_j$ set $p_j = p_{j-1}$ and $\delta_k = \widetilde{\delta}_j$. Otherwise, continue.
10:     Set $j = j+1$.
11: **end while**
12: Set $k = k+1$ and return to 2.

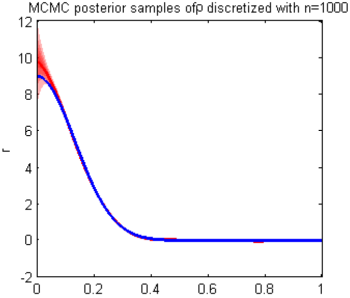
Marginalized δ chains

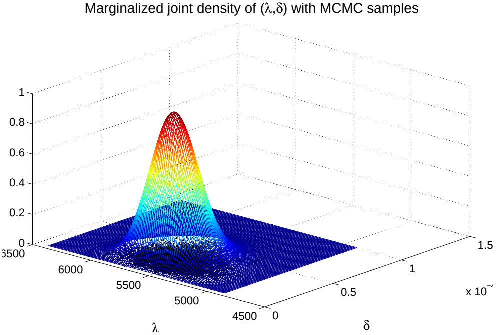


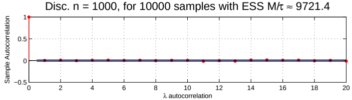
Marginalized δ chains

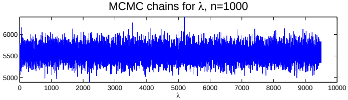


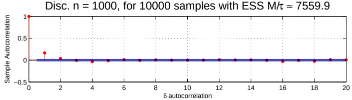
Results

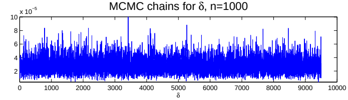

MCMC posterior samples ofρ discretized with n=1000
r


Marginalized joint density of (λ,δ) with MCMC samples
λ
δ


Disc. n = 1000, for 10000 samples with ESS M/τ ≈ 9721.4
Sample Autocorrelation
λ autocorrelation


MCMC chains for λ, n=1000
λ


Disc. n = 1000, for 10000 samples with ESS M/τ ≈ 7559.9
Sample Autocorrelation
δ autocorrelation


MCMC chains for δ, n=1000
δ

Future Work

- The current implementation takes approximately 5-6 hours on a Intel(R) Core(TM) i5-4300U CPU @ 1.90GHz core for a reconstruction with $n = 1000$. Other factorizations based on spectral methods might allow for faster than $O(n^3)$ computation for calucating U and c in each Metropolis step.
- Analyze how uncertainty in the blurring kernel affects deconvolution uncertainty.
- Theoretical details related to the infinite dimensional prior $\mathcal{L}\rho$ are not fully developed.
- Explore other priors that might enforce higher correlation near the peak of the PSFs.

References

Sergios Agapiou, Johnathan M Bardsley, Omiros Papaspiliopoulos, and Andrew M Stuart.
Analysis of the Gibbs sampler for hierarchical inverse problems.
SIAM/ASA Journal on Uncertainty Quantification, 2(1):511–544, 2014.

Johnathan M Bardsley.
MCMC-based image reconstruction with uncertainty quantification.
SIAM Journal on Scientific Computing, 34(3):A1316–A1332, 2012.

Daniela Calveti and Erkki Somersalo.
Hypermodels in the Bayesian imaging framework.
Inverse Problems, 24(3):034013, 208.

Andrew M Stuart.
The Bayesian approach to inverse problems.
arXiv preprint arXiv:1302.6989v1, 2013.