Point Spread Function Estimation and Uncertainty Quantification

Kevin Joyce[†*]
advised by John Bardsley[†] and Aaron Luttman[*]

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

May 5, 2016

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.

Outline

Modeling Imaging Systems

- Convolution with a point spread function
- Estimating the PSF with calibration images

Radial Symmetry for Function Spaces

- Sobolev Spaces
- Variable Transformation and the Pullback Operator
- Regularization and Discrete representation

Hierarchical Bayesian Model

- The posterior density
- Gibbs Sampling and Partial Collapse
- Results

Outline

Modeling Imaging Systems
- Convolution with a point spread function
- Estimating the PSF with calibration images

Radial Symmetry for Function Spaces
- Sobolev Spaces
- Variable Transformation and the Pullback Operator
- Regularization and Discrete representation

Hierarchical Bayesian Model
- The posterior density
- Gibbs Sampling and Partial Collapse
- Results

Imaging Model Assumptions

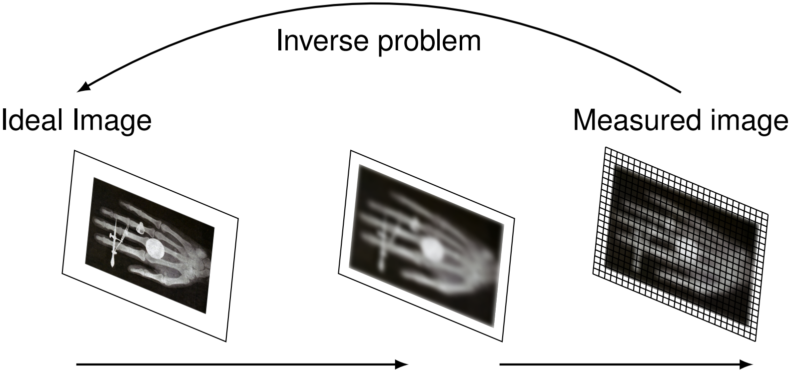


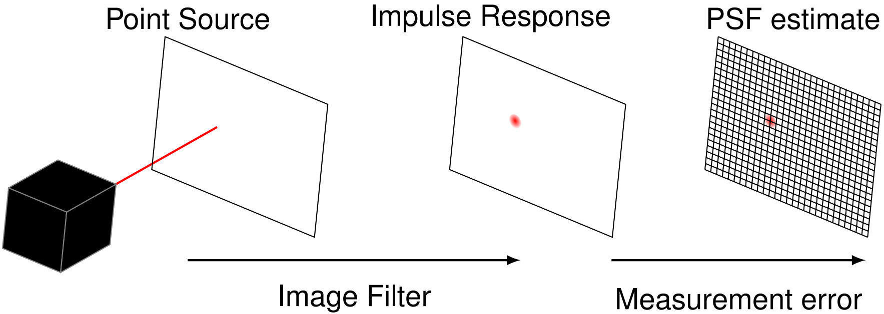
Image filter

$$b(x) = \iint p(|x - s|) \cdot f(s)\, ds$$

Measurement error

$$+\varepsilon \sim \text{white noise}$$

Point Spread Function Estimation

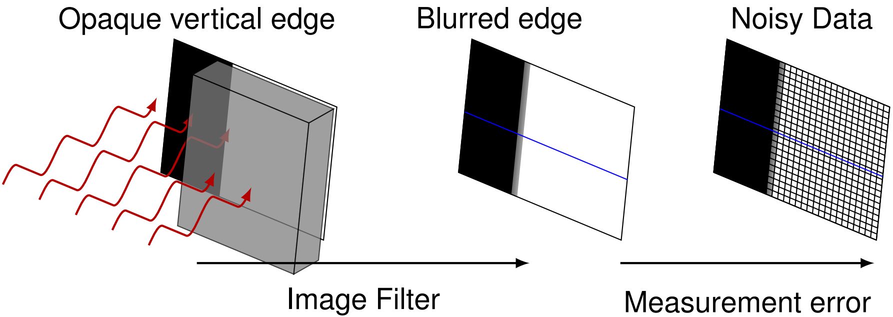


$$\iint \color{darkred}{p(|x-s|)}\,\delta(s)\,ds$$

$$\iint \color{darkred}{p(|s|)}\,\delta(x-s)\,ds = \color{darkred}{p(|x|)}$$

$+\varepsilon \sim$ white noise

Point Spread Function Estimation

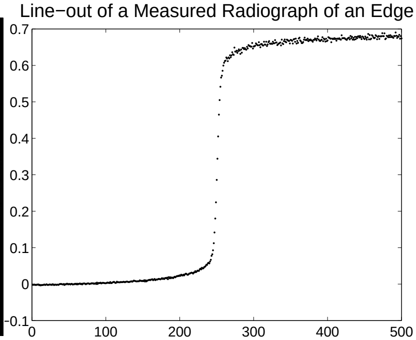


$$\iint p(|x-s|)E(s)\,ds$$

$$\iint p(|s|)E(x-s)\,ds = b(x)$$

$+\varepsilon \sim$ white noise

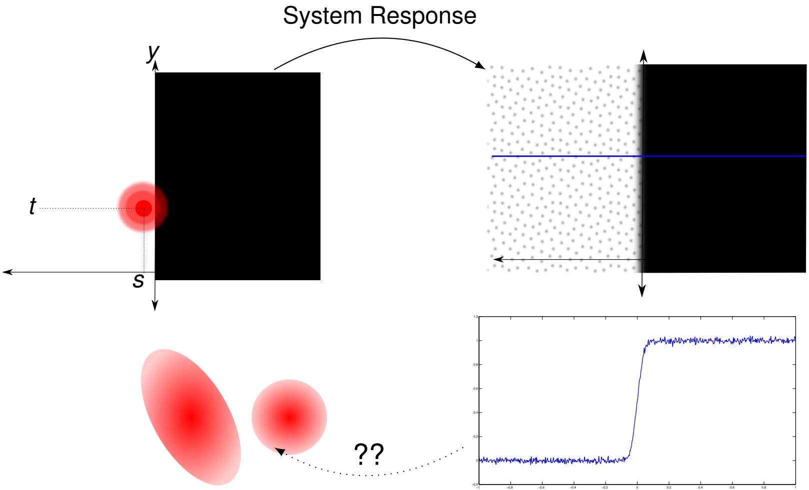
X-ray Edge Calibration Data



Radiographic data from the Cygnus Dual Beam Radiography Facility at the NNSS in North Las Vegas.

General Edge Blur Problem

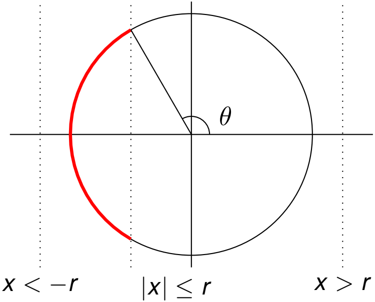
$$b(x,y) = \iint_{\mathbb{R}^2} k(s,t) E(x-s)\, dt\, ds + \varepsilon_{x,y}, \quad E(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0. \end{cases}$$

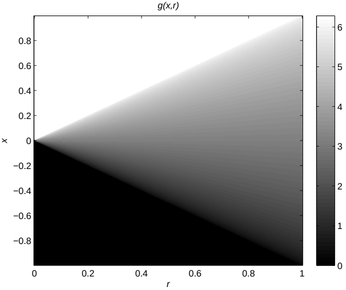


Radially Symmetric PSF

We distinguish the radial profile from the kernel by
$k(s,t) = p\left(\sqrt{s^2+t^2}\right)$

$$b(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} k(s,t)E(x-s)dtds + \varepsilon_{x,y}$$
$$= \int_0^{\infty} p(r)\cdot g(x,r)\, rdr + \varepsilon_{x,y}.$$





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

Outline

Distributions and Sobolev spaces

- Let $\phi \in \mathscr{D}(\Omega)$ denote the space of compactly supported smooth functions defined on an open set $\Omega \subseteq \mathbb{R}^N$, called test functions.

- The space of continuous linear functionals, denoted $f \in \mathscr{D}^*(\Omega)$, are the distributions on Ω, where action of f on ϕ is expressed by $\langle f, \phi \rangle$.

- For functions, the action of the linear functional is $\langle f, g \rangle = \int fg \, dx$.

- Operations are expressed adjointly, e.g. differentiation is given by integration by parts $Df(\phi) \stackrel{\text{def}}{=} -\langle f, D\phi \rangle$.

Distributions and Sobolev spaces

- We define the L^2 inner-product for test functions as the sesquilinear form $(\cdot,\cdot)_{L^2(\Omega)} : \mathscr{D}(\Omega) \times \mathscr{D}(\Omega) \to \mathbb{C}$ by the Riemann integral

  $$(\phi, \psi)_{L^2(\Omega)} \stackrel{\text{def}}{=} \int_\Omega \phi(x)\overline{\psi(x)}\, dx,$$

  with a norm $\| \cdot \|_{L^2(\Omega)}$.

- We can construct L^2 from test functions with a completion argument. Idea: Equivalence classes of L^2 Cauchy sequences of test functions (ϕ_n) correspond to L^2 distributions.

Distributions and Sobolev spaces

- A Sobolev space of order n over an open set $\Omega \subseteq \mathbb{R}^k$ is $\mathscr{H}^n(\Omega) = \{f \in L^2(\Omega) : \partial^\alpha f \in L^2(\Omega) \text{ whenever } |\alpha| \leq N\}$.
- They are endowed with the sum of semi-norms

$$(f, g)_{\Omega,n} = \sum_{0 \leq |\alpha| \leq n} (\partial^\alpha f, \partial^\alpha g)_{L^2(\Omega)} \,. \tag{1}$$

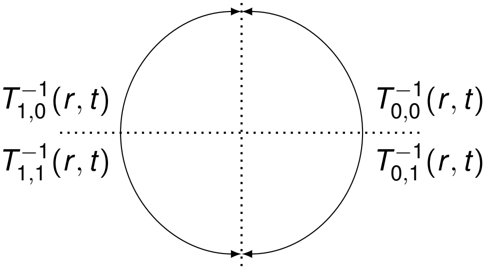
- Each of these form a sequence of linear subspaces $\mathscr{H}^n(\Omega) \subset \mathscr{H}^{n-1}(\Omega) \subset \cdots \subset \mathscr{H}^1(\Omega) \subset L^2(\Omega)$, however, the inclusion is strict and they are not closed with respect to the L^2 norm.

Variable Transformation and the Pullback Operator

- Idea: Extend the notion of $k(x, y) = p\left(\sqrt{x^2+y^2}\right) = T^\sharp p$ to distributions "adjointly" as was done for derivatives: $\langle T^\sharp p, \phi\rangle \stackrel{\text{def}}{=} \langle p, T_\sharp \phi\rangle$

- Use change of variables so that one component is $r = \sqrt{x^2+y^2}$, for topological reasons, this can only be done on a proper subset of $\mathbb{R}^2$.

- Let $T_{ij}(x, y) = \left(\sqrt{x^2+y^2}, (-1)^j y\right)$,



Radial Symmetry for Sobolev Spaces

- The pullback by T on $\mathscr{D}^*(\Omega_1)$ is a linear operator $T^\sharp : \mathscr{D}^*(\Omega_1) \to \mathscr{D}^*(\Omega_2)$ that is injective, continuous, and unique.

- **Definition** $k \in \mathscr{K}^n \subset \mathscr{H}^n(\Omega_2)$, the space of radially symmetric distributions, if there exists a sequence $(\rho_m) \subset \mathscr{D}(\Omega_1)$, so that $(T^\sharp \rho_m)$ is Cauchy with respect to $\|\cdot\|_{\mathscr{H}^k(\Omega_2)}$ and

$$\left\langle k, \phi \right\rangle_{\Omega_2} = \lim_{n\to\infty} \left\langle T^\sharp \rho_n, \phi \right\rangle_{\Omega_2} = \lim_{m\to\infty} \left\langle \rho_m, T_\sharp \phi \right\rangle_{\Omega_1},$$

- **Definition** The space of radial profiles corresponding to $\mathscr{K}^n$ distributions is $\mathscr{P}^n = \{p \in \mathscr{D}^*(\Omega_1) : T^\sharp p \in \mathscr{K}^n\}$.

Radial Symmetry for Sobolev Spaces

- The map $T^\sharp$ induces the inner product

$$(\rho, \omega)_{T(\Omega_1)} = \Big(S_{1/2}(\rho), S_{1/2}(\omega)\Big)_{L^2(\Omega_1)}$$

  where $S(\omega)$ is the shift operator defined by
  $S(\omega) = \omega(r) \cdot (2\pi r)^{1/2}$.

- When k is a function, the familiar radial transformation is given

$$\iint |k|^2 \, dxdy = \int |p|^2 \, 2\pi r dr.$$

- Moreover, if $\rho, \omega \in \mathscr{D}(\Omega_1)$, then the squared norm of the Laplacian is given by

$$(\nabla T^\sharp \rho, \nabla T^\sharp \omega)_{L^2(\Omega_2)} = (\partial\rho, \partial\omega)_{T(\Omega_1)}.$$

Radial L^2 and the Laplacian

For the two representations $k = T^\sharp p$

$$b = \mathcal{F}k \quad \Longrightarrow \quad b = \left(\mathcal{F}T^\sharp\right) p = \mathcal{G}p$$

- When k and p are smooth, real-valued functions, the Laplacian on $\mathbb{R}^2$ with the L^2 inner product translates to

$$\begin{aligned}\|\nabla k\|^2_{L^2(\Omega_2)} = (\partial p, \partial p)_{T(\Omega_1)} &= \int_0^\infty \frac{d}{dr}p(r) \cdot \frac{d}{dr}p(r) r dr \\ &= \int_0^\infty p \cdot \underbrace{\frac{1}{r}\frac{d}{dr} r \frac{d}{dr}}_{\mathcal{L}} p(r) r dr\end{aligned}$$

- We solve the inverse problem on the radial profile p, with regularization on the radially symmetric function k.

Tikhonov Laplacian Regularization

For the two representations

$$b = \mathcal{G}p + \epsilon \quad \text{and} \quad b = \mathcal{F}k + \epsilon$$

- Minimizing the second order Tikhonov-Laplacian functional subject to k radially symmetric

$$\frac{\lambda}{2}\left\|b - \mathcal{G}p\right\|_{L^2}^2 + \frac{\delta}{2}\left\|\nabla^2 k\right\|_{L^2}^2$$

- is equivalent to minimizing

$$\frac{\lambda}{2}\left\|b - \mathcal{G}p\right\|_{L^2}^2 + \frac{\delta}{2}\left\|\mathcal{L}^2 p\right\|_{rad}^2$$

The discrete problem

In order to carry out estimation on a computer, we discretize the integral operator using mid-point quadrature

$$b = \mathcal{G}p + \epsilon \quad \Longrightarrow \quad \boldsymbol{b} = \boldsymbol{G}\boldsymbol{p} + \boldsymbol{\epsilon}$$

Further, we discretize the regularization operator $\mathcal{L}$ using finite differencing

$$\|\mathcal{L}p\|^2_{rad} = \int \left[\frac{1}{r}\frac{d}{dr}\left(r\frac{d}{dr}\right)\right]^2 p(r) r dr \;\Longrightarrow\; \boldsymbol{L}\boldsymbol{p} = \boldsymbol{r}^{-1/2} \odot \boldsymbol{D}\left(\boldsymbol{r} \odot \boldsymbol{D}\boldsymbol{p}\right)$$

and midpoint quadrature for the inner products

$$\frac{\lambda}{2}\left\|b - \mathcal{G}p\right\|^2_{L^2} \Longrightarrow \underbrace{\frac{\lambda}{2m}}_{\lambda}\left\|\boldsymbol{b} - \boldsymbol{G}\boldsymbol{p}\right\|^2_{\mathbb{R}^m}$$

and

$$\frac{\delta}{2}\left\|\nabla^2 k\right\|_{L^2} \Longrightarrow \frac{\delta}{2}\left\|\mathcal{L}p\right\|^2_{rad} \Longrightarrow \underbrace{\frac{\delta}{2n}}_{\delta}\left\|\boldsymbol{L}\boldsymbol{p}\right\|^2_{\mathbb{R}^n}$$

Outline

Hierarchical Model for PSF estimation

Let $\pi(\boldsymbol{x}) = \mathbb{P}(X = x)$ denote the probability density. For λ, δ and $\boldsymbol{p}$ and

$$\boldsymbol{b} = \boldsymbol{G}\boldsymbol{p} + \epsilon$$

assume

- The likelihood $\pi(\boldsymbol{b}|\boldsymbol{p}, \lambda, \delta) = \pi(\boldsymbol{b}|\boldsymbol{p}, \lambda) \propto \lambda^{M/2} \exp\left(-\frac{\lambda}{2}\|\boldsymbol{b} - \boldsymbol{G}\boldsymbol{p}\|^2\right)$ since ϵ is independent Gaussian noise.
- The prior $\pi(\boldsymbol{p}|\delta, \lambda) = \pi(\boldsymbol{p}|\delta) \propto \delta^{N/2} \exp\left(-\frac{\delta}{2}\|\boldsymbol{L}\boldsymbol{p}\|^2\right)$ since $k \sim \mathcal{N}(0, \nabla^{-2}) \implies p \sim N(0, \mathcal{L}^{-2})$
- The hyperpriors $\pi(\lambda) \propto \exp\left(-10^{-4}\lambda\right)$ and $\pi(\delta) \propto \exp\left(-10^{-4}\delta\right)$ are independent “unobjective” Gamma distributions.

Bayesian Posterior

With $\pi(\boldsymbol{b}|\boldsymbol{p},\lambda,\delta), \pi(\boldsymbol{p}|\delta,\lambda)$, and $\pi(\lambda,\delta)$, use Bayes' "Theorem" to obtain

$$\pi_{\mathbf{b}}(\boldsymbol{p},\lambda,\delta) \stackrel{\text{def}}{=} \pi(\boldsymbol{p},\lambda,\delta|\boldsymbol{b}) = \pi(\boldsymbol{b},\boldsymbol{p},\lambda,\delta)/\pi(\boldsymbol{b})$$
$$\propto \lambda^{M/2}\delta^{N/2}\exp\left(-\frac{\lambda}{2}\left\|\boldsymbol{b}-\boldsymbol{G}\boldsymbol{p}\right\|_{\mathbb{R}^m} - \frac{\delta}{2}\left\|\boldsymbol{L}\boldsymbol{p}\right\|_{\mathbb{R}^n} - 10^{-4}(\lambda+\delta)\right)$$

- This is not a "common" probability density, hence simulations from a computer are not readily available.

- Bayes' "Theorem" will allow simulations from the full conditionals $\pi_{\mathbf{b}}(\lambda|\delta,\boldsymbol{p}), \pi_{\mathbf{b}}(\delta|\lambda,\boldsymbol{p})$ and $\pi_{\mathbf{b}}(\boldsymbol{p}|\lambda,\delta)$.

- Because each distribution is from the exponential family, they form a conjugacy such that the full conditionals are "shifts" of the priors.

Full conditional densities

- The resulting expressions are

$$\pi(\lambda|\boldsymbol{b},\boldsymbol{p},\delta) \propto \lambda^{(2N+1)/2+\alpha-1} \exp\left(-\lambda\left(\frac{1}{2}\|\boldsymbol{Gx}-\boldsymbol{b}\|^2-\beta\right)\right),$$

$$\pi(\delta|\boldsymbol{b},\boldsymbol{p},\lambda) \propto \delta^{N/2+\alpha-1} \exp\left(-\delta\left(\frac{1}{2}\langle\boldsymbol{p},\boldsymbol{Lp}\rangle-\beta\right)\right),$$

$$\pi(\boldsymbol{p}|\boldsymbol{b},\lambda,\delta) \propto \exp\left(-\frac{1}{2}\left\langle(\boldsymbol{p}-\boldsymbol{m}_{\lambda,\delta}),\boldsymbol{J}_{\lambda,\delta}(\boldsymbol{p}-\boldsymbol{m}_{\lambda,\delta})\right\rangle\right)$$

where

$$\boldsymbol{J}_{\lambda,\delta} \stackrel{\text{def}}{=} (\lambda\boldsymbol{G}^T\boldsymbol{G}+\delta\boldsymbol{L}) \quad \text{and} \quad \boldsymbol{m}_{\lambda,\delta} \stackrel{\text{def}}{=} \boldsymbol{J}_{\lambda,\delta}^{-1}\lambda\boldsymbol{G}^T\boldsymbol{b},$$

- The matrix solves required for sampling can be efficiently computed using a Cholesky decomposition $\boldsymbol{R}_{\lambda,\delta}^T\boldsymbol{R}_{\lambda,\delta} \stackrel{\text{def}}{=} \boldsymbol{J}_{\lambda,\delta}$ in $O(N^3)$ flops.

Gibbs sampling

The Gibbs sampler [Geman and Geman 1984]:
Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}\right)$, simulate

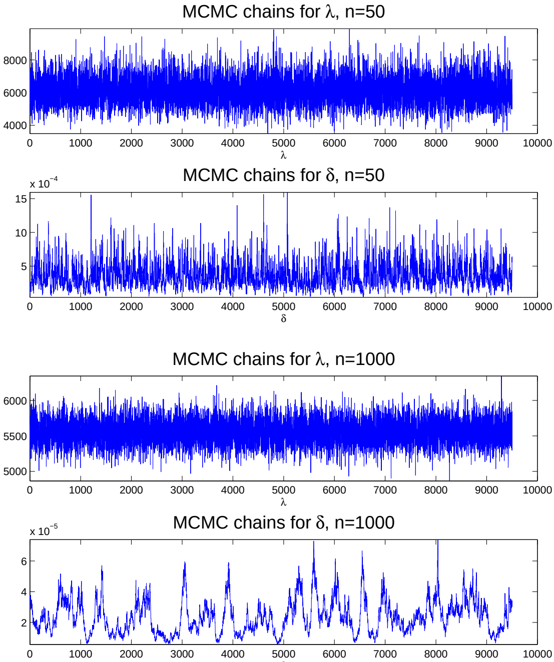
1. Simulate $\lambda^{k+1} \sim \Gamma\left((2N+1)/2 + \alpha, \frac{1}{2}\|\boldsymbol{G}\boldsymbol{p}^k - \boldsymbol{b}\|^2 + \beta\right)$.

2. Simulate $\delta^{k+1} \sim \Gamma\left(N/2 + \alpha, \frac{1}{2}\left\langle \boldsymbol{p}^k, \boldsymbol{L}\boldsymbol{p}^k \right\rangle + \beta\right)$.

3. Compute $\boldsymbol{R}_{\lambda^{k+1},\delta^{k+1}}$, $\boldsymbol{m}_{\lambda^{k+1},\delta^{k+1}}$,

   and set $\boldsymbol{p}^{k+1} = \boldsymbol{R}^{-1}_{\lambda^{k+1},\delta^{k+1}}\boldsymbol{z} + \boldsymbol{m}_{\lambda^{k+1},\delta^{k+1}}$ where $\boldsymbol{z} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_{N\times N}\right)$.

Correlated δ chains



Literature on the Issue

- [Agapiou,Bardsley,Stuart,Papaspiliopoulos, 2014] explained this phenomena theoretically for a general class of Laplacian based Hierarchical samplers for inverse problems.
- The issue arises when the discretization of $\boldsymbol{p}$ closer approximates the continuum, the correlation in the δ component of the Markov Chain becomes more correlated.
- [VanDyke, Park 2008] provide a general method for removing the dependence of problematic components in the Gibbs sampler, called partial collapse.
- The idea has been independently derived in many places, however, if done carelessly [VanDyke, Park 2008] showed that the resulting Markov chain is no longer invariant, although invariance was not proved there.

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda|\delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}}|\lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p}|\lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}|\lambda, \delta)$:

$$
\begin{aligned}
[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] &= \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda')\pi_{\mathbf{b}}(\lambda', \boldsymbol{p}|\delta)\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}} \\
&= \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \iint \pi_{\mathbf{b}}(\lambda'|\delta, \boldsymbol{p}) \underbrace{\int\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\widetilde{\boldsymbol{p}}d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p} \\
&= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda')/\pi_{\mathbf{b}}(\delta', \lambda')} \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p})d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')} \\
&= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta')\frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')} \\
&= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')
\end{aligned}
$$

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda|\delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}}|\lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p}|\lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}|\lambda, \delta)$:

$$[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] = \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda')\pi_{\mathbf{b}}(\lambda', \boldsymbol{p}|\delta)\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \iint \pi_{\mathbf{b}}(\lambda'|\delta, \boldsymbol{p}) \underbrace{\int\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\widetilde{\boldsymbol{p}} d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p}$$

$$= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda')/\pi_{\mathbf{b}}(\delta', \lambda')} \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p})d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta')\frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')}$$

$$= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')$$

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda|\delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}}|\lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p}|\lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}|\lambda, \delta)$:

$$[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] = \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda')\pi_{\mathbf{b}}(\lambda', \boldsymbol{p}|\delta)\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \iint \pi_{\mathbf{b}}(\lambda'|\delta, \boldsymbol{p}) \underbrace{\int\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\widetilde{\boldsymbol{p}} d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p}$$

$$= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda')/\pi_{\mathbf{b}}(\delta', \lambda')} \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p})d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta')\frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')}$$

$$= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')$$

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda | \delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}} | \lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p} | \lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p}) \pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}} | \lambda, \delta)$:

$$[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] = \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}' | \lambda', \delta') \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}' | \lambda') \pi_{\mathbf{b}}(\lambda', \boldsymbol{p} | \delta) \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}' | \lambda', \delta') \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}' | \lambda') \iint \pi_{\mathbf{b}}(\lambda' | \delta, \boldsymbol{p}) \underbrace{\int \int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) d\widetilde{\boldsymbol{p}} d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p}$$

$$= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}' | \lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda') / \pi_{\mathbf{b}}(\delta', \lambda')} \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}' | \lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p}) d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta') \frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')}$$

$$= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')$$

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda|\delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}}|\lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p}|\lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}|\lambda, \delta)$:

$$[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] = \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda')\pi_{\mathbf{b}}(\lambda', \boldsymbol{p}|\delta)\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \iint \pi_{\mathbf{b}}(\lambda'|\delta, \boldsymbol{p}) \underbrace{\int\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\widetilde{\boldsymbol{p}}d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p}$$

$$= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda')/\pi_{\mathbf{b}}(\delta', \lambda')} \quad \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p})d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')}$$

$$= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta')\frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')}$$

$$= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')$$

Marginalized Sampler

Given $\left(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1}, \widetilde{\boldsymbol{p}}^{k-1}\right)$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda|\delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $(\delta^k, \widetilde{\boldsymbol{p}}^k) \sim \pi_{\mathbf{b}}(\delta, \widetilde{\boldsymbol{p}}|\lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p}|\lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to $\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}|\lambda, \delta)$:

$$
\begin{aligned}
[\mathcal{K}\widetilde{\pi_{\mathbf{b}}}] &= \iiiint \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda')\pi_{\mathbf{b}}(\lambda', \boldsymbol{p}|\delta)\widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\lambda d\delta d\boldsymbol{p} d\widetilde{\boldsymbol{p}} \\
&= \pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')\pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \iint \pi_{\mathbf{b}}(\lambda'|\delta, \boldsymbol{p}) \underbrace{\int\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}})d\widetilde{\boldsymbol{p}}d\lambda}_{\pi_{\mathbf{b}}(\delta, \mathbf{p})} d\delta d\boldsymbol{p} \\
&= \underbrace{\pi_{\mathbf{b}}(\boldsymbol{p}'|\lambda', \delta')}_{\pi_{\mathbf{b}}(\mathbf{p}', \delta', \lambda')/\pi_{\mathbf{b}}(\delta', \lambda')} \pi_{\mathbf{b}}(\delta', \widetilde{\boldsymbol{p}}'|\lambda') \underbrace{\iint \pi_{\mathbf{b}}(\lambda', \delta, \boldsymbol{p})d\delta d\boldsymbol{p}}_{\pi_{\mathbf{b}}(\lambda')} \\
&= \pi_{\mathbf{b}}(\boldsymbol{p}', \lambda', \delta')\frac{\pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}}', \lambda', \delta')}{\pi_{\mathbf{b}}(\lambda', \delta')} \\
&= \widetilde{\pi_{\mathbf{b}}}(\lambda', \delta', \boldsymbol{p}', \widetilde{\boldsymbol{p}}')
\end{aligned}
$$

Partially Collapsed Sampler

Given $(\lambda^{k-1}, \delta^{k-1}, \boldsymbol{p}^{k-1})$, simulate

1: $\lambda^k \sim \pi_{\mathbf{b}}(\lambda | \delta^{k-1}, \boldsymbol{p}^{k-1})$
2: $\delta^k \sim \pi_{\mathbf{b}}(\delta | \lambda^k)$
3: $\boldsymbol{p}^k \sim \pi_{\mathbf{b}}(\boldsymbol{p} | \lambda^k, \delta^k)$

The associated Markov Chain is invariant with respect to
$\int \widetilde{\pi_{\mathbf{b}}}(\lambda, \delta, \boldsymbol{p}, \widetilde{\boldsymbol{p}}) = \int \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p}) \pi_{\mathbf{b}}(\widetilde{\boldsymbol{p}} | \lambda, \delta) = \pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})$.

- The order of the chain matters in the previous arguments.
- Permuting steps 2 and 3 results in a chain that is no longer invariant with respect to $\pi_{\mathbf{b}}$.
- Cyclically permuting results in a different sampler as well, however, this does not practically effect the overall chain, only the first and last steps, e.g. $(\mathcal{K}_1 \mathcal{K}_3 \mathcal{K}_2)^N = \mathcal{K}_1 (\mathcal{K}_3 \mathcal{K}_2 \mathcal{K}_1)^{N-1} \mathcal{K}_2 \mathcal{K}_3$

Marginalized Posterior Density

In order to sample $\pi_{\mathbf{b}}(\delta|\lambda)$, we complete the square of the quadratic form in $\pi_{\mathbf{b}}(\lambda, \delta, \boldsymbol{p})$ and integrate out λ, this results in

$$\pi_{\mathbf{b}}(\delta|\lambda) \propto \exp\left((n/2)\delta - \ln|\det \boldsymbol{J}_{\lambda,\delta}| - \frac{\lambda}{2}\langle \boldsymbol{b}, \boldsymbol{H}_{\lambda,\delta}\boldsymbol{b}\rangle - 10^{-4}\delta \right),$$

where $\boldsymbol{J}_{\lambda,\delta} = \lambda \boldsymbol{G}^T\boldsymbol{G} + \delta \boldsymbol{L}$ and $\boldsymbol{H}_{\lambda,\delta} = \boldsymbol{I} - \lambda \boldsymbol{G}\boldsymbol{J}_{\lambda,\delta}^{-1}\boldsymbol{G}^T$.

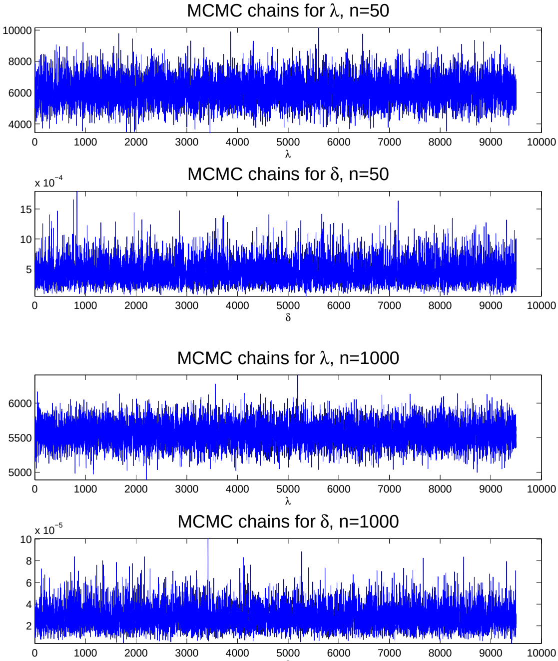
- To draw from this density, we embed a Metropolis-Hastings algorithm within the Gibbs sampler.
- It can be shown that the resulting MCMC algorithm remains invariant.
- **Corrollary** Suppose $\mathcal{K} = \mathcal{K}_m \ldots \mathcal{K}_1$ is the transition operator for the Gibbs sampler, and $\widetilde{\mathcal{K}}_i$ given $\boldsymbol{x}_{\widehat{i}}$ is an operator such that $\widetilde{\mathcal{K}}_i\left[\pi(x_i|\boldsymbol{x}_{\widehat{i}})\right] = \pi(x_1'|\boldsymbol{x}_{\widehat{i}})$, then $\mathcal{K}_m \ldots \widetilde{\mathcal{K}}_i\mathcal{K}_{i-1} \ldots \mathcal{K}$ is invariant with respect to π.

Metropolis-Hastings within Gibbs Sampling

- We use n_{MH} Metropolis-Hastings steps using a Gaussian proposal with variance γ.
- Due to numerical overflow issues, all computations are carried out on the log scale.
- The full implementation is:

1: Let λ_0, δ_0,$\boldsymbol{p}$ and γ be given.
2: Draw λ_{k+1} from $\Gamma\left(n/2+\alpha,\ \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}_{\boldsymbol{k}}-\boldsymbol{b}\|^2-\beta\right)$.
3: Set $j=1$, Compute $\boldsymbol{R}_0^T\boldsymbol{R}_0=\boldsymbol{J}_{\lambda^k,\delta^{k-1}}$, then $\pi_0=\log\pi_{\mathbf{b}}(\delta^{k-1}|\lambda^k)$.
4: **for** $1<j<n_{MH}$ **do**
5: Draw $\widetilde{\delta}$ from $N(\widetilde{\delta}_{j-1},\gamma)$.
6: Compute $\widetilde{\boldsymbol{R}}^T\widetilde{\boldsymbol{R}}=J_{\lambda^k,\widetilde{\delta}}$, then $\pi_j=\log p(\widetilde{\delta}_j|\boldsymbol{f}\lambda_k)$.
7: Set $\widetilde{\pi}=\pi_j$, $\boldsymbol{R}_j=\widetilde{\boldsymbol{R}}$ and $\delta_k=\widetilde{\delta}_j$ with probability $\widetilde{\pi}/\pi_j$
8: **end for**
9: Draw $\boldsymbol{p}_k$ from $N\left(\boldsymbol{R}_k^{-1}\lambda_k\boldsymbol{A}^T\boldsymbol{b},\boldsymbol{R}_k^{-1}\right)$.
10: Set $k=k+1$ and return to 2.

Marginalized δ chains



Integrated Autocorrelation

- The variance of the chain-mean estimator for the mean of the invariant density π is

$$\mathrm{Var}(\overline{X}_N) = \frac{1}{N^2}\sum_{k=1}^{N}\mathrm{Var}(X^k) + \frac{1}{N^2}\sum_{k\neq l}^{N}\mathrm{Cov}(X^l, X^k).$$

- When (X_k) are assumed to be nearly identically distributed via the ergodic theorem, then for large N, we can approximate

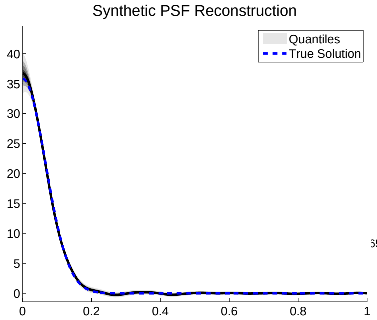
$$\mathrm{Var}(\overline{X}_N) \approx \frac{\sigma^2}{N}\sum_{k=-\infty}^{\infty}\rho(k) \text{ where } \rho(k) \stackrel{\text{def}}{=} \frac{\mathrm{Cov}(X^1, X^{|k|})}{\sigma^2}.$$

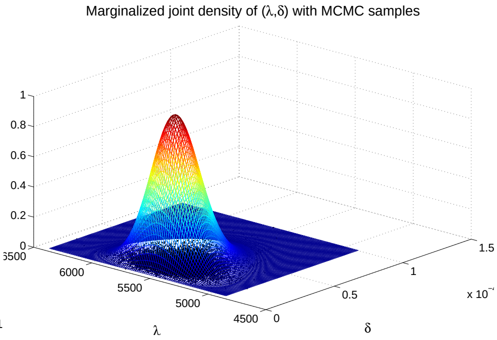
- The function ρ is called the normalized auto-correlation function and the parameter $\tau_{\text{int}} \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty}\rho(k)$ is the integrated auto-correlation time.

Comparing algorithms

-

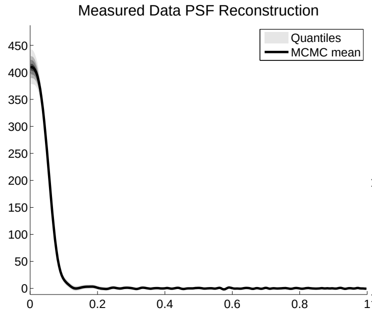
Results

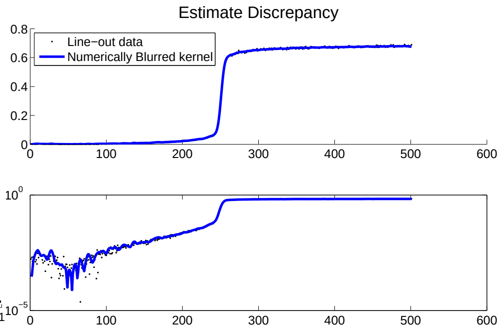




Algorithm	$\widehat{\lambda}_{\text{MCMC}}$ $(\times 10^4)$	$\widehat{\delta}_{\text{MCMC}}$ $(\times 10^{-8})$	λ-p_{Geweke}	δ-p_{Geweke}	IACT	ESS	#Chol/ESS
Gibbs	1.102	6.132	0.998	0.850	36.2	138.0	72.4
PC Gibbs $n_{mh} = 1$	1.102	5.611	0.992	0.943	7.9	633.0	31.6
PC Gibbs $n_{mh} = 5$	1.102	5.515	0.999	0.985	1.3	3799.6	15.8

Results





Algorithm	$\widehat{\lambda}_{\mathrm{MCMC}}$ $(\times 10^4)$	$\widehat{\delta}_{\mathrm{MCMC}}$ $(\times 10^{-10})$	λ-p_{Geweke}	δ-p_{Geweke}	IACT	ESS	#Chol/ESS
Gibbs	9.146	1.245	0.995	0.964	14.0	357.6	28.0
PC Gibbs $n_{mh} = 1$	9.167	1.191	0.995	0.998	8.5	587.3	34.1
PC Gibbs $n_{mh} = 5$	9.178	1.189	0.994	0.980	1.5	3278.5	18.3
MTC	9.090	1.200	0.996	0.969	12.5	432.2	23.1

Summary and Future Work

- We introduced a novel Hierarchical Bayesian non-parametric model for estimating translation invariant and isotropic image blur with and edge.
- We developed the Partially Collapsed Gibbs sampler from the Gibbs sampler, and showed when partial collapse remained stationary.
- We then implemented the algorithm on a synthetic example using Metropolis with Partially Collapsed Gibbs, and showed that it improves the standard Gibbs sampler.
- Future: Develop the model and algorithm completely in infinite dimensions.
- Future: Adapt the strategies to other imaging models that incorporate radial geometry such as Abel and Radon transforms.

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 Calvetti and Erkki Somersalo.
An Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing, volume 2.
Springer Science & Business Media, 2007.

A. M. Stuart.
Inverse problems: A Bayesian perspective.
Acta Numerica, 19:451–559, 2010.

David A Van Dyk and Taeyoung Park.
Partially collapsed Gibbs samplers: Theory and methods.
Journal of the American Statistical Association, 103(482):790–796, 2008.