

Bayesian Density Estimation via Multiple Sequential Inversions of 2-D Images with Application in Electron Microscopy

Dalia Chakrabarty^{*,||} Fabio Rigat^{†,**} Nare Gabrielyan^{‡,††} Richard Beanland^{§,‡‡} and Shashi Paul^{¶,††} ,

|| Department of Statistics

University of Warwick

Coventry CV4 7AL, U.K.

e-mail: d.chakrabarty@warwick.ac.uk

and

Department of Mathematics

University of Leicester

Leicester LE1 7RH, U.K.

e-mail: dc252@le.ac.uk

*** Novartis Vaccines and Diagnostics*

via Fiorentina 1, Siena, Italy

fabio.rigat@novartis.com

†† Emerging Technology Research Centre

DeMontfort University

Hawthorn Building, The Gateway

Leicester, LE1 9BH

U.K.

nare.gabrielyan@email.dmu.ac.uk

spaul@dmu.ac.uk

‡‡ Department of Physics

University of Warwick

Coventry CV4 7AL, U.K.

e-mail: r.beanland@warwick.co.uk

Throughout, we refer to our main manuscript as CRGBP.

^{*} Associate Research fellow at Department of Statistics, University of Warwick and Lecturer of Statistics at Department of Mathematics, University of Leicester

[†] Research Biostatistics Group Head, Novartis Vaccines and Diagnostics and Associate fellow at Department of Statistics, University of Warwick

[‡] Graduate student at Emerging Technologies Research Centre, De Montfort University

[§] Lecturer, Department of Physics, University of Warwick

[¶] Head, Emerging Technologies Research Centre, De Montfort University

1. Identifying low-rank and spatially-varying components of density

It is known in the literature that the general, under-determined deconvolution problem is solvable only if the unknown density is intrinsically, “sufficiently” sparse (Donoho & Tanner 2005; Wright et al. 2009). Here we advance a methodology to learn the density - sparse or dense, , within the frame of the designed experiment discussed above.

The density is function recognized to be made of a constant ρ_0 (the low-rank component in the limiting sense), and the spatially varying component $\rho_1(x, y, z)$ that may be sparse or dense in \mathbb{R}^3 . We view the constant part of the density as $\rho_0 = \rho_0 \delta(x - x_i, y - y_i, z)$ where $\delta(\cdot, \cdot, \cdot)$ is the Dirac delta function on \mathbb{R}^3 (Chakraborty 2008), centered at the center of the ik -th interaction volume, $\forall k = 1, 2, \dots, N_{\text{eng}}$. Then in our problem, the contribution of the constant part of the density to the projection onto the center of the ik -th interaction-volume is

$$\begin{aligned} \mathcal{C}(\rho_0 * \eta(x, y, z))_i^{(k)} &\equiv \rho_0 \mathcal{C}(\delta(x - x_i, y - y_i, z) * \eta(x, y, z))_i^{(k)} \\ &= I0^{(k)}, \end{aligned} \quad (1.1)$$

a constant independent of the beam pointing location i , if $\eta(x, y, z)$ is restricted to be a function of the depth coordinate Z only. As is discussed in Section 4 of CRGBP, this is indeed what we adopt in the model for the kernel.

Then, $I0^{(k)}$ depends only on the known morphological details of the interaction-volume for a given value of E , $\forall i = 1, 2, \dots, N_{\text{data}}$. Thus, $\{\tilde{I}_i^{(k)}\}_{i=1}^{N_{\text{data}}} = \{\tilde{I}_i^{(k)} + I0^{(k)}\}_{i=1}^{N_{\text{data}}}$, where $\tilde{I}_i^{(k)}$ is the spatially-varying component of the image data. The identification of the constant component of the density is easily performed as due to the constant component of the measurable.

In our inversion exercise, it is the $\{\tilde{I}_i^{(k)}\}_{i=1}^{N_{\text{data}}}$ field that is actually implemented as data, after $I0^{(k)} := \inf\{\tilde{I}_i^{(k)}\}_{i=1}^{N_{\text{data}}}$ is subtracted from $\{\tilde{I}_i^{(k)}\}_{i=1}^{N_{\text{data}}}$, for each $k = 1, \dots, N_{\text{eng}}$. Hereafter, when we refer to the data, the spatially-varying part of the data will be implied; it is this part of the data that will hereafter be referred to as $\{\tilde{I}_i^{(k)}\}_{i=1}^{N_{\text{data}}}$, at each value ϵ_k of E , $k = 1, \dots, N_{\text{eng}}$. Its inversion will yield a spatially varying sparse/dense density, that we will from now, refer to as $\rho(x, y, z)$ that in general lies in a non-convex subset of $\mathbb{R}_{\geq 0}$. Thus, we see that in this model, it is possible for $\rho(x, y, z)$ to be 0. The construction of the full density, inclusive of the low-rank and spatially-varying parts, is straightforward once the latter is learnt.

2. Discretized versions of sequential projection

In this section we present the discretized form of Equation 3.4 of CRGBP, where this equation presents the sequential projection of the convolution $\rho * \eta$ of the unknown material density $\rho(x, y, z)$ and microscopy correction function $\eta(z)$, onto the center of the general (ik -th) interaction-volume. Here the beam incidence location index $i = 1, \dots, N_{\text{data}}$ and the energy of the electrons of the incident beam takes the value ϵ_k , with $k = 1, \dots, N_{\text{eng}}$. Equation 3.4 of CRGBP takes different forms depending on the resolution in the image data; we work with 3 different resolution classes where the resolution ω of the data determines the size of the square cross-sectional area of a voxel at any depth under the surface of the material sample.

2.1. 1st model: high- Z systems, coarse resolution

When dealing with “high- Z ” materials (see Equation 3.6 of CRGBP for definition), imaged at the coarsest resolution available, we recall that the material density inside an interaction-volume at a given value of the sub-surface depth Z , is the density inside a single voxel, at that Z . Then the material density inside an interaction-volume at a given Z is isotropic (see Section 3.1.1 of CRGBP). Then recalling that the discrete convolution $\rho * \eta$ within the k -th Z -bin and at the i -th beam pointing gives $(\rho * \eta)_i^{(k)} = \sum_{m=1}^k \xi_i^{(m)} \eta^{(k-m)}$, we get the projection $\mathcal{C}(\rho * \eta)_i^{(k)}$ of the convolution onto the center of the ik -th interaction-volume to be the discretized form of Equation 3.4 of CRGBP:

$$\mathcal{C}(\rho * \eta)_i^{(k)} = \frac{1}{(R0^{(k)})^2} \sum_{q=0}^k \left[\frac{(R0^{(q)})^2 - (R0^{(q-1)})^2}{2} \left\{ \sum_{t=0}^q \left(h^{(t)} - h^{(t-1)} \right) \sum_{m=1}^t \xi_i^{(m)} \eta^{(t-m)} \right\} \right]. \quad (2.1)$$

2.2. 2nd model: low- Z systems, coarse resolution

In this case, the material density inside the ik -th interaction volume is isotropic for $k = 1, 2, \dots, k_{in}$, i.e. at any Z , the ik -th interaction volume is confined to a given voxel. However, at higher values of beam energies, namely, for $k = k_{in+1}, \dots, N_{eng}$, the ik -th interaction volume spills over into the neighboring voxels, at a given Z . Let the ik' -th voxel lie wholly inside the ik -th interaction-volume and let the $i'k'$ -th voxel be its neighbor. Let the fraction of the volume of the $i'k'$ -th voxel, contained within the ik -th interaction volume be $w_{i'|i}^{(k')}$, where $k' \leq k$.

In general, at a given Z , any bulk voxel has 8 neighboring voxels and when the voxel lies at the corner or edge of the sample, number of nearest neighbors is less than 8. Then, at a given Z , there will be contribution from at most 9 voxels towards $\mathcal{C}(\rho * \eta)_i^{(k)}$. At $Z = z \in [h^{(k'-1)}, h^{(k')}]$, for any i , let the maximum number of contributing nearest neighbors be $i_{max}|i, k$ so that $i_{max}|i, k \leq 9$. The notation for this number bears its dependence on both i and k . We define $\bar{\xi}_i^{(k')}$ as the weighted average of the densities in the ik' -th voxel and its nearest neighbors that are fully or partially included within the ik -th interaction-volume. Here $k' \leq k$, $k = 1, \dots, N_{eng}$, $i = 1, \dots, N_{data}$. Thus,

$$\bar{\xi}_i^{(k')} := \sum_{i'=1}^{i_{max}|i, k'} \xi_{i'|i}^{(k')} w_{i'|i}^{(k')}, \quad (2.2)$$

where the i' -th neighbor of the ik' -th voxel at the same depth, harbors the density $\xi_{i'|i}^{(k')}$ and there is a maximum of $i_{max}|i, k'$ such neighbors. The effect of this averaging over the nearest neighbors at this depth, is equivalent to averaging over the angular coordinate θ and results in the angular averaged density $\bar{\xi}_i^{(k')}$ at this Z , which by definition, is isotropic, i.e. independent of the angular coordinate. Then for $k > k_{in}$, $\mathcal{C}(\rho * \eta)_i^{(k)}$ is computed as in Equation 2.1 with $\xi_i^{(\cdot)}$ on the RHS of this equation replaced by the isotropic angular averaged density $\bar{\xi}_i^{(\cdot)}$. However, for $k \leq k_{in}$, the projection is computed as in Equation 2.1.

2.3. 3rd model: fine resolution

In this class of resolution, the resolution is so fine, i.e. ω is so small that $\omega \ll R0^{(N_{\text{eng}})}$. For this model, the projection equation is written in terms of the Cartesian coordinates (x, y, z) of a point instead of the polar coordinate representation of this point, where the point in question lies inside the ik -th interaction-volume that is centered at $(x_i, y_i, 0)$. Then inside the ik -th interaction-volume, at a given x and y , $z \in \left[0, \sqrt{(R0^{(k)})^2 - (x - x_i)^2 - (y - y_i)^2}\right]$. For

$$x - x_i \in [(u - 1)\omega, u\omega] \quad u = -(\text{int})\left(\frac{R0^{(k)}}{\omega}\right) + 1, -(\text{int})\left(\frac{R0^{(k)}}{\omega}\right) + 2, \dots, (\text{int})\left(\frac{R0^{(k)}}{\omega}\right),$$

the index $p_u(k)$ of the Y -bin of voxels lying fully inside the ik -th interaction volume, with respect to the center of this interaction-volume, are

$$p_u(k) = -q_u(k), -q_u(k) + 1, \dots, 0, 1, 2, \dots, q_u(k) - 1, q_u(k),$$

where

$$q_u(k) := (\text{int})\left(\frac{\sqrt{(R0^{(k)})^2 - u^2\omega^2}}{\omega}\right).$$

Then using the definition of the beam-pointing index in terms of the X -bin and Y -bin indices of voxels (see Equation 3.1 of CRGBP), we get the beam-pointing index $\varrho_u(i, k)$ of voxels lying wholly inside the ik -th interaction-volume, for a given u is

$$\varrho_u(i, k) = i - q_u(k)\sqrt{N_{\text{data}}} + u, i - (q_u(k) - 1)\sqrt{N_{\text{data}}} + u, \dots, i - (q_u(k) - 2q_u(k))\sqrt{N_{\text{data}}} + u,$$

i.e. for a given u , $\varrho_u(i, k) = i + p_u(k)\sqrt{N_{\text{data}}} + u$.

The depth coordinate of voxels with beam-pointing index $\varrho_u(i, k)$ lying inside the ik -th interaction-volume are $z \in \left[0, \sqrt{(R0^{(k)})^2 - (p_u(k))^2\omega^2 - u^2\omega^2}\right]$ so that the energy index of voxels lying fully inside at Y -bin $p_u(k)$ and $x - x_i \in [(u - 1)\omega, u\omega]$ are $\in [1, t_{\text{max}}(u)]$ where $t_{\text{max}}(u) \in \mathbb{Z}_{>0}$ such that $t_{\text{max}}(u) = \max\{1, 2, \dots, N_{\text{eng}}\}$ that satisfies

$$h^{(t_{\text{max}}(u))} \leq \sqrt{(R0^{(k)})^2 - (p_u(k))^2\omega^2 - u^2\omega^2}.$$

At this Y -bin index $p_u(k)$, there will also exist a voxel lying partly inside the ik -th interaction-volume, at the $(t_{\text{max}}(u) + 1)$ -th Z -bin, between depths $h^{t_{\text{max}}(u)}$ and $\sqrt{(R0^{(k)})^2 - (p_u(k))^2\omega^2 - u^2\omega^2}$. In addition, the projection $\mathcal{C}(\rho \star \eta)_i^{(k)}$ will include contributions from voxels at the edge of this interaction-volume, lying partly inside it; the beam-pointing indices of such voxels will be $i - (q_u(k) + 1)\sqrt{N_{\text{data}}} + u$ and $i + (q_u(k) + 1)\sqrt{N_{\text{data}}} + u$ for $x - x_i \in [(u - 1)\omega, u\omega]$ with u and $q(u)$ defined as above. Lastly, parts of voxels at beam-pointing indices $i - (\text{int})\left(\frac{R0^{(k)}}{\omega}\right) - 1$ and $i + (\text{int})\left(\frac{R0^{(k)}}{\omega}\right) + 1$ will also be contained inside the ik -th interaction-volume. These voxels at the edges extend into the 1st Z -bin. We can compute the fraction $r_a^{(b)}(i, k)$ of the volume of the ab -th voxel contained partly within the ik -th interaction-volume by tracking the geometry of the system. Then using the discretized version of Equation 3.4 of CRGBP, we write,

$$\begin{aligned}
\omega^{-2}(R0^{(k)})^2 \mathcal{C}(\rho * \eta)_i^{(k)} = & \\
& \sum_{u=-(int)(R0^{(k)}/\omega)}^{(int)(R0^{(k)}/\omega)} \sum_{p_u(k)=-q_u(k)}^{q_u(k)} \sum_{t=1}^{t_{max}(u)} \left[\left(h^{(t)} - h^{(t-1)} \right) \sum_{m=1}^t \xi_{\varrho_u(i,k)}^{(m)} \eta^{(t-m)} \right] + \\
& \sum_{u=-(int)(R0^{(k)}/\omega)}^{(int)(R0^{(k)}/\omega)} \sum_{p_u(k)=-q_u(k)}^{q_u(k)} \left[\left(\sqrt{(R0^{(k)})^2 - ((q_u(k))^2 + u^2)\omega^2} - h^{(t_{max}(u))} \right) \sum_{m=1}^{t_{max}(u)+1} \chi_{\varrho_u}^{(m)}(i,k) \eta^{(t_{max}(u)+1-m)} \right] + \\
& \sum_{\ell(i,k)} \left[\left(h^{(1)} \right) r_{\ell}^{(1)}(i,k) \xi_{\ell(i,k)}^{(1)} \eta^{(0)} \right]
\end{aligned} \tag{2.3}$$

where $\ell(i,k) = i - (int) \left(\frac{R0^{(k)}}{\omega} \right) - 1, i + (int) \left(\frac{R0^{(k)}}{\omega} \right) + 1, i - (q_u(k) + 1) \sqrt{N_{\text{data}}} + u, i + (q_u(k) + 1) \sqrt{N_{\text{data}}} + u$, for $u = -(int) \left(\frac{R0^{(k)}}{\omega} \right) + 1, -(int) \left(\frac{R0^{(k)}}{\omega} \right) + 2, \dots, (int) \left(\frac{R0^{(k)}}{\omega} \right)$,
 $\chi_{\varrho_u}^{(m)}(i,k) := r_{\varrho_u}^{(m)}(i,k) \xi_{\varrho_u(i,k)}^{(m)}$,
and $\eta^{(1)}$ is the surface ionization i.e. the measured value of the kernel on the system surface (see Section 4 of CRGBP).

3. Priors on sparsity: illustrations on simulated densities

In order to illustrate that the prior probability density developed in Section 5 of CRGBP, sensitively adapts to the sparsity in the material density distribution, we present results of 2 simulation studies (shown in Figure 1). In these studies, the density parameter values in the ik -th voxel are simulated from 2 simplistic toy models that differ from each other in the degree of sparsity of the true material density distribution: $\xi_i^{(k)} = u_1^{10}/u_2$, and $\xi_i^{(k)} = u_3^{10}$ respectively, (where u_1, u_2, u_3 are uniformly distributed random numbers in $[0, 1]$), at a chosen i and energy indices $k = 1, 2, \dots, 10$. In the simulations we specify the beam penetration depth $h^{(k)} \propto \epsilon_k^{1.67}$ as suggested by Kanaya & Okamaya (1972); as any interaction-volume is hemispherical, its radius $R0^{(k)} = h^{(k)}$. The kernel parameters $\eta^{(k)}$ are generated from a quadratic function of $h^{(k)}$ with noise added. In the simulations, the material is imaged at resolution ω such that $\pi[R0^{(10)}]^2 \leq \omega^2$, i.e. the “1st model” is relevant (see Section 3.3.1 of CRGBP). This allows for simplification of the computation of $\mathcal{C}(\rho * \eta)_i^{(k)}$ according to Equation 3.4 of CRGBP. Then at this i , for $k = 1, 2, \dots, 10$, $\xi_i^{(k)}$ are plotted in Figure 1 against k , as is the logarithm of the prior $\pi_0(\xi_i^{(k)})$ computed according to Equation 5.4 of CRGBP, with p held as a random number, uniform in $[0.6, 0.99]$. Logarithm of the priors are also plotted as a function of the material density parameter. We see from the figure that the prior developed here tracks the sparsity of the vector $(\xi_i^{(1)}, \xi_i^{(2)}, \dots, \xi_i^{(10)})^T$ well.

4. Inference

In this work, we learn the unknown material density and kernel parameters using the mismatch between the data $\{\tilde{I}_i^{(k)}\}_{k=1; i=1}^{k=N_{\text{eng}}; i=N_{\text{data}}}$ and $\{\mathcal{C}(\rho * \eta)_i^{(k)}\}_{k=1; i=1}^{k=N_{\text{eng}}; i=N_{\text{data}}}$, in terms of which, the likelihood is defined. The material density and kernel are convolved, and this convolution is sequentially projected onto the center of the the ik -th interaction volume, in the model (out of the 3 models, depending on the resolution of the image data at hand).

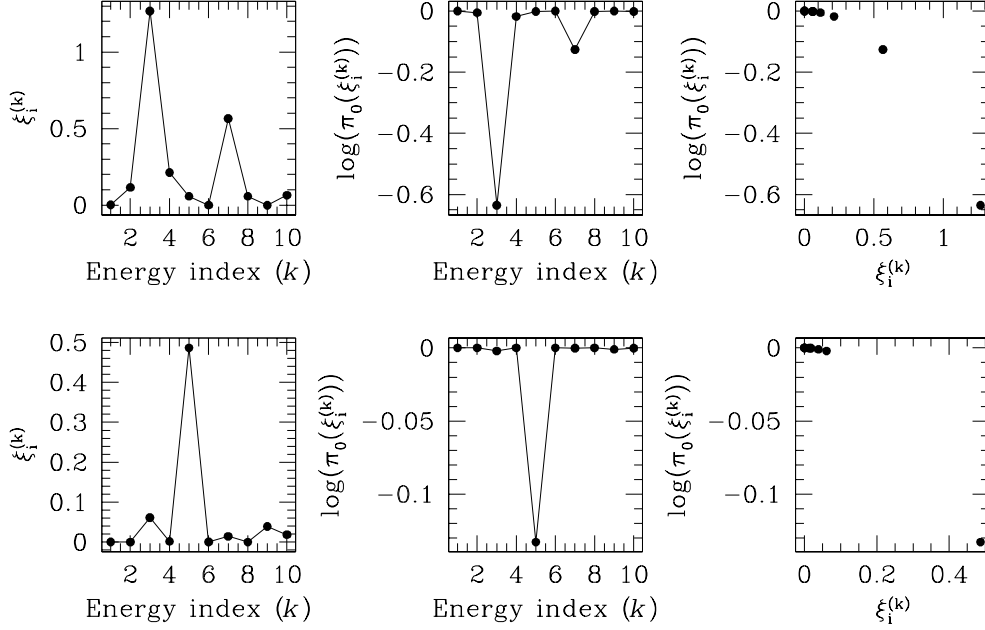


FIG 1. *Top: in the left panel black filled circles depict values of simulated material density parameters $\xi_i^{(k)} = u_1^{10}/u_2$, $u_1, u_2 \sim \mathcal{U}[0, 1]$, at an arbitrary beam position index i , as a function of the energy index k , for $k = 1, 2, \dots, 10$. Log of the prior $\pi_0(\xi_i^{(k)})$, as given in Equation 5.4 of CRGBP, is shown in the middle panel as a function of k for $p \sim \mathcal{U}[0.6, 0.99]$. The log prior is plotted against the true values of $\xi_i^{(k)}$ in black filled circles in the right panel. Bottom: As in the top panels, except that this simulation is of a sparser material density distribution with density parameters generated as $\xi_i^{(k)} = u_1^{10}$.*

We choose to work with a Gaussian likelihood:

$$\mathcal{L} \left(\xi_1^{(1)}, \dots, \xi_1^{(N_{\text{eng}})}, \dots, \xi_{N_{\text{data}}}^{(1)}, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}, \eta^{(1)}, \dots, \eta^{(N_{\text{eng}})} | \tilde{I}_1^{(1)}, \tilde{I}_1^{(2)}, \dots, \tilde{I}_1^{(N_{\text{eng}})}, \tilde{I}_2^{(1)}, \dots, \tilde{I}_2^{(N_{\text{eng}})}, \dots, \tilde{I}_{N_{\text{data}}}^{(N_{\text{eng}})} \right) = \prod_{k=1}^{N_{\text{eng}}} \prod_{i=1}^{N_{\text{data}}} \frac{1}{\sqrt{2\pi}\sigma_i^{(k)}} \exp \left[-\frac{\left(\mathcal{C}(\rho * \eta)_i^{(k)} - \tilde{I}_i^{(k)} \right)^2}{2 \left(\sigma_i^{(k)} \right)^2} \right], \quad (4.1)$$

where the noise in the image datum $\tilde{I}_i^{(k)}$ is $\sigma_i^{(k)}$; it is discussed in Section 3.4 of CRGBP.

Towards the learning of the unknown functions, the joint posterior probability density of the unknown parameters, given the image data, is $\pi \left(\xi_1^{(1)}, \dots, \xi_1^{(N_{\text{eng}})}, \dots, \xi_{N_{\text{data}}}^{(1)}, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}, \eta^{(1)}, \dots, \eta^{(N_{\text{eng}})} | \tilde{I}_1^{(1)}, \dots, \tilde{I}_{N_{\text{data}}}^{(N_{\text{eng}})} \right)$, defined using Bayes rule in terms of the likelihood (Equation 4.1), the adaptive prior probability on the sparsity of the density function (Section 5 of CRGBP) and the prior on the kernel, (Section 4 of CRGBP). Once the posterior probability density of the material density function and kernel, given the image data is defined, we use the adaptive Metropolis within Gibbs (Haario, Laine, Mira & Saksman 2006) to generate posterior samples.

At the n -th iteration, $n = 1, \dots, N_{\text{max}}$, $\xi_i^{(k)}$ is proposed from a folded normal density¹. This choice of the proposal density is motivated by a non-zero probability for $\xi_i^{(k)}$ to be zero. The latter constraint rules out a gamma or

beta density that $\xi_i^{(k)}$ is proposed from but truncated and folded normal densities are acceptable; $k = 1 \dots, N_{\text{eng}}$, $i = 1, \dots, N_{\text{data}}$. Of these we choose the easily computable folded normal proposal density (Leone, Nottingham & Nelson 1961). The proposed density in the n -th iteration, in the ik -th voxel is

$$\tilde{\xi}_i^{(k)}|_n \sim \mathcal{N}_F(\mu_i^{(k)}|_n, \varsigma_i^{(k)}|_n) \quad (4.2)$$

while the current density in this voxel at the n -th iteration is defined as $\xi_i^{(k)}|_n$. We choose the mean and variance of this proposal density to be

$$\begin{aligned} \mu_i^{(k)}|_n &= \xi_i^{(k)}|_{n-1}, \quad \forall n = 1, \dots, N_{\text{max}} \\ \left(\varsigma_i^{(k)}|_n\right)^2 &= \begin{cases} \frac{\sum_{p=n_0}^{n-1} \left(\xi_i^{(k)}|_p\right)^2}{n - n_0} - \left[\frac{\sum_{p=n_0}^{n-1} \left(\xi_i^{(k)}|_p\right)}{n - n_0} \right]^2 & \text{if } n \geq n_0 \\ T\xi_i^{(k)}|_0 & \text{if } n < n_0 \end{cases} \end{aligned} \quad (4.3)$$

The random variable T is considered to be uniformly distributed, i.e. $T \sim U(0, 1]$. Thus, for $n \geq n_0$, the proposal density is adaptive, (Haario et al. 2006). We choose $n_0 = 10^3$ and N_{max} is of the order of 8×10^4 .

We choose $\xi_i^{(k)}|_0$ by assigning constant density to the voxels that constitute the ik -th interaction-volume, $k = 1 \dots, N_{\text{eng}}$, $i = 1, \dots, N_{\text{data}}$.

When a distribution-free model for the kernel is used, in the n -th iteration, $\eta^{(k)}$ is proposed from a folded normal proposal density with mean set by the current value of this parameter and an experimentally chosen dispersion s_1 . When the parametric model for the kernel is used, $\eta(z)$ is calculated as given in Equation 4.2 of CRGBP, conditional on the values of 2 the parameters Q and η_0 . The proposed parameters at the n -th iteration are \tilde{Q}_n and $(\tilde{\eta}_0)_n$. \tilde{Q}_n and $(\tilde{\eta}_0)_n$ are each proposed from independent exponential proposal densities with constant rate parameters.

Inference is performed by sampling from the high dimensional posterior using Metropolis-within-Gibbs block update, (Gilks & Roberts 1996; Chib & Greenberg 1995). Let the state vector at the n -th iteration be

$$\varepsilon_n = (\xi_1^{(1)}|_n, \dots, \xi_1^{(N_{\text{eng}})}|_n, \dots, \xi_2^{(N_{\text{eng}})}|_n, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}|_n, \eta^{(1)}|_n, \dots, \eta^{(N_{\text{eng}})}|_n)^T. \quad (4.4)$$

For the implementation of the block Metropolis-Hastings, we partition the state vector ε_n as:

$$\varepsilon_n^T = ((\varepsilon_n^{(\xi)})^T, (\varepsilon_n^{(\eta)})^T),$$

where

$$\begin{aligned} \varepsilon_n^{(\xi)} &= (\xi_1^{(1)}|_n, \dots, \xi_1^{(N_{\text{eng}})}|_n, \dots, \xi_2^{(N_{\text{eng}})}|_n, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}|_n)^T, \\ \varepsilon_n^{(\eta)} &= (\eta^{(1)}|_n, \dots, \eta^{(N_{\text{eng}})}|_n)^T. \end{aligned} \quad (4.5)$$

Here $n = 1, \dots, N_{\text{burnin}}, \dots, N_{\text{max}}$. We typically use $N_{\text{max}} > 8 \times 10^4$ and $N_{\text{burnin}} = 5 \times 10^3$. Then, the state ε_{n+1} is given by the successive updating of the two blocks: $\varepsilon_{n+1}^{(\xi)}$ and $\varepsilon_{n+1}^{(\eta)}$.

¹The distribution $\mathcal{N}_F(a, b)$ is the folded normal distribution with mean $a \in \mathbb{R}$, $a > 0$ and standard deviation $b \in \mathbb{R}$, $b > 0$ (Leone et al. 1961)

5. In the small noise limit

Theorem 5.1. *In the limit of small noise, $\sigma_i^{(k)} \rightarrow 0$, the joint posterior probability of the density and kernel parameters, given the image data, for all beam-pointing indices ($i = 1, \dots, N_{\text{data}}$) and all ϵ_k , $k = 1, \dots, N_{\text{eng}}$, reduces to a product of $N_{\text{data}} \times N_{\text{eng}}$ Dirac measures, with the ik -th measure centered at the solution to the equation $\tilde{I}_i^{(k)} = \mathcal{C}(\rho * \eta)_i^{(k)}$,*

Proof. Recalling the developed priors on the density parameters and on the kernel parameters and the Gaussian likelihood, logarithm of the posterior probability of the discretized distribution-free model is

$$\begin{aligned} \log \pi \left(\xi_1^{(1)}, \dots, \xi_1^{(N_{\text{eng}})}, \dots, \xi_{N_{\text{data}}}^{(1)}, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}, \eta^{(1)}, \dots, \eta^{(N_{\text{eng}})} | \tilde{I}_1^{(1)}, \dots, \tilde{I}_{N_{\text{data}}}^{(N_{\text{eng}})} \right) = \\ \sum_{i=1}^{N_{\text{data}}} \sum_{k=1}^{N_{\text{eng}}} \left[-\log \sigma_i^{(k)} - \left(\frac{(\tilde{I}_i^{(k)} - \mathcal{C}(\rho * \eta)_i^{(k)})^2}{2(\sigma_i^{(k)})^2} \right) \right] - \sum_{k=1}^{N_{\text{eng}}} \left[\frac{(\eta^{(k)} + \eta_0^{(k)})^2}{2N(s^{(k)})^2} \right] - \sum_{i=1}^{N_{\text{data}}} \sum_{k=1}^{N_{\text{eng}}} \left[\left(\xi_i^{(k)} \nu(\tau_i^{(k)}) \right)^2 \right] + A, \end{aligned} \quad (5.1)$$

where $A \in \mathbb{R}$ is a finite constant. Thus,

$$\begin{aligned} \lim_{\sigma_i^{(k)} \rightarrow 0} \pi(\xi_1^{(1)}, \dots, \xi_{N_{\text{data}}}^{(N_{\text{eng}})}, \eta^{(1)}, \dots, \eta^{(m)} | \tilde{I}_1^{(1)}, \dots, \tilde{I}_{N_{\text{data}}}^{(N_{\text{eng}})}) \propto \\ \lim_{\sigma_i^{(k)} \rightarrow 0} \left[\prod_{i=1}^{N_{\text{data}}} \prod_{k=1}^{N_{\text{eng}}} \frac{1}{\sigma_i^{(k)}} \exp \left(-\frac{(\tilde{I}_i^{(k)} - \mathcal{C}(\rho * \eta)_i^{(k)})^2}{2(\sigma_i^{(k)})^2} \right) \right]. \end{aligned} \quad (5.2)$$

The right hand side of this equation is the product of Dirac delta functions centered at $\tilde{I}_i^{(k)} = \mathcal{C}(\rho * \eta)_i^{(k)}$, for $i = 1, \dots, N_{\text{data}}$, $k = 1, \dots, N_{\text{eng}}$. Thus, the joint posterior probability density of the unknowns reduces to a product of Dirac measures for each i, k , with each measure centered on the solution of the equation $\tilde{I}_i^{(k)} = \mathcal{C}(\rho * \eta)_i^{(k)}$. \square

References

- Chakrabarty, S. (2008), “Some Applications of Dirac’s Delta Function in Statistics for More Than One Random Variable,” *Applications and Applied Mathematics*, 3(1), 42–54.
- Chib, S., & Greenberg, E. (1995), “Understanding the Metropolis-Hastings Algorithm,” *The American Statistician*, 49(4), 327.
- Donoho, D., & Tanner, J. (2005), “Sparse Nonnegative Solutions of Underdetermined Linear Equations by Linear Programming,” *Proceedings of the National Academy of Sciences*, 102(27), 9446.
- Gilks, W. R., & Roberts, G. O. (1996), Strategies for improving MCMC,, in *Markov Chain Monte Carlo in Practice*, eds. W. Gilks, S. Richardson, & D. Spiegelhalter, Interdisciplinary Statistics, Chapman and Hall, London, pp. 89–114.
- Haario, H., Laine, M., Mira, A., & Saksman, E. (2006), “DRAM: Efficient adaptive MCMC,” *Statistics and Computing*, 16, 339.

- Kanaya, K., & Okamaya, S. (1972) *Jl. of Physics D., Applied Physics*, 5, 43.
- Leone, F., Nottingham, R. B., & Nelson, L. S. (1961), “The Folded Normal Distribution,” *Technometrics*, 3(4), 543.
- Wright, J., Yang, A., Ganesh, A., Sastry, S., & Ma, Y. (2009), “Robust Face Recognition via Sparse Representation, and its online supplementary material,” *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 31(2).