Bayesian Denoising/Deblurring of Poisson-Gaussian Corrupted Data Using Complex Wavelets

Timothy D. Roberts and Nick Kingsbury
Sigproc Laboratory, Cambridge University Engineering Department (CUED), Cambridge, U.K.

***Abstract*—In this work we summarize our algorithm used to denoise/deblur observations which result from a mixture of Poisson and Gaussian noise sources. Our algorithm, MSIST-P, is an extension of the MSIST algorithm, in that it accounts for a spatially varying statistical distribution. Using our own simulated ground truth data, we've obtained better reconstruction results than the Richardson Lucy algorithm.**

I. Notations

In this document we use the following matrix/vector notations:

- $\mathbf{x}$ is a vector corresponding to the ground-truth image;
- $\mathbf{A}$ is a matrix representing the forward blurring operator;
- $\mathbf{A}^T$ is a matrix representing the inverse blurring operator;
- $\mathbf{b}$ is a constant vector representing the background signal;
- $\mathbf{y}$ is a random vector modeling the measurements;
- $\mathbf{y}_b$ is simply $\mathbf{y} - \mathbf{b}$;
- $\mathbf{x}_n$ is an estimate of $\mathbf{x}$ obtained after n iterations of the algorithm;
- $\mathbf{w}$ is a j-indexed with real and imaginary parts in alternate locations, comprising the coefficients of a complex-valued wavelet representation;
- $\mathbf{w}_n$ is the wavelet transform of $\mathbf{x}_n$ obtained after n iterations of the algorithm;
- $\mathbf{w}_y$ is the wavelet transform of $\mathbf{y}$;
- $\mathbf{W}$ is a real matrix representing a forward wavelet transformation;
- $\mathbf{W}^T$ is a real matrix representing an inverse wavelet transformation;
- ν^2 is a regularization parameter representing the assumed variance of the additive Gaussian noise;
- ϵ^2 is a regularization parameter which controls the shape of the re-weighted L2 sparse penalty $\mathbf{w}^T\mathbf{S}\mathbf{w}$;
- ν^2 is a regularization parameter which controls the relative strength of the re-weighted L2 sparse penalty $\mathbf{w}^T\mathbf{S}\mathbf{w}$;
- $\mathbf{\Lambda}_\alpha$ is a diagonal matrix of wavelet subband weights, which approximates $\mathbf{W}\mathbf{A}^T\mathbf{A}\mathbf{W}^T$;
- $\mathbf{\Phi}$ is $\mathbf{A}\mathbf{W}^T$;
- $\sigma^2_{w,j}$ is the variance of the j-th coefficient in $\mathbf{w}$;
- $\mathbf{S}$ is a j-indexed diagonal matrix of wavelet coefficient inverse variance estimates;
- s_j is the j-element of $\mathbf{S}$;
- $\mathbf{\Psi_y}$ is a j-indexed real diagonal matrix of scaling coefficients corresponding to $\mathbf{w}_y$;
- $\mathbf{\Sigma}_w$ is a diagonal matrix whose elements are the elements of column vector $\nu^2\mathbf{1} + \mathbf{\Psi_y}$;
- $\mathcal{T}\{\mathbf{q}, \mathbf{b}\}$ is a thresholding operator which sets elements of vector $\mathbf{q}$ less than $\mathbf{b}$ to 0, this represents the noise variance of $\mathbf{w}_y$.
- $\mathbf{g} \sim N(0, \sigma_g^2)$ is an additive white Gaussian noise vector (detector noise);
- $\mathcal{Q}\{\mathbf{q}\}$ denotes the quantization of $\mathbf{q}$ to 16-bit unsigned integer precision;
- $\mathcal{P}(\boldsymbol{\lambda})$ denotes a Poisson random vector with mean $\boldsymbol{\lambda}$, which models the shot noise of our imaging system.

II. Algorithm

The MSIST-P algorithm follows an extension of the basic MSIST algorithm [1] and applies the Gaussian scale mixture model (GSM) [3] to a combination of Gaussian and Poisson noise sources. It involves simple matrix-vector multiplies and indexing operations for each iteration:

$$\mathbf{w}_{n+1} = (\mathbf{\Lambda}_\alpha + \mathbf{\Sigma}_w\mathbf{S}_n)^{-1}[(\mathbf{\Lambda}_\alpha - \mathbf{\Phi}^T\mathbf{\Phi})\mathbf{z}_n + \mathbf{\Phi}^T\mathbf{y}_b] \quad (1)$$

$$\mathbf{z}_{n+1} = \mathbf{w}_{n+1} \quad (2)$$

$$\mathbf{x}_{n+1} = \mathcal{T}\left\{\mathbf{W}^T\mathbf{w}_{n+1}, \mathbf{0}\right\} \quad (3)$$

$$S_{2j,n+1} = S_{2j-1,n+1} = \frac{1}{\sigma^2_{w,j}} = \frac{1}{\frac{1}{2}(w^2_{2j,n} + w^2_{2j-1,n}) + \epsilon^2} \quad (4)$$

This represents a threshold Landweber iteration in the wavelet domain, followed by a hard-thresholding in the spatial domain, at each iteration. In this algorithm, we have chosen to use the 3D version of the DT-$\mathbb{C}$WT for $\mathbf{W}$ because of it's directional selectivity and approximate shift-invariance properties.

III. Variational interpretation

The MSIST-P algorithm assumes that the following observation model can be simplified by assuming a single Gaussian noise source with a spatially-varying variance.

$$\mathbf{y} = \mathcal{Q}\left\{\mathcal{P}(\mathbf{A}\mathbf{x} + \mathbf{b}) + \mathbf{g}\right\} \quad (5)$$

We develop a cost function and algorithm here in order to recover $\mathbf{x}$ from these blurred, noisy, quantized measurements.

Neglecting $Q\{\}$, we can rewrite eq. (5) using our spatially-varying Gaussian assumption:

$$\mathbf{y} \sim N(\boldsymbol{\lambda}, \mathbf{\Sigma}) \quad (6)$$

where $\boldsymbol{\lambda} = \mathbf{A}\mathbf{x} + \mathbf{b}$ is the local variance due to the Poisson noise, and $\boldsymbol{\Sigma}$ is a diagonal matrix comprised of the elements of $\boldsymbol{\sigma}^2 = \boldsymbol{\lambda} + \sigma_g^2 \mathbf{1}$.

Thus, eq. (6) results in the following posterior:

$$\mathrm{p}(\mathbf{w}|\mathbf{y}) \propto \mathrm{p}(\mathbf{y}|\mathbf{w})\mathrm{p}(\mathbf{w}|\mathbf{S})p(\mathbf{S}) \tag{7}$$

where we assume that $\boldsymbol{\lambda} = \mathbf{A}\mathbf{x} + \mathbf{b} = \mathbf{A}\mathbf{W}^T\mathbf{w} + \mathbf{b}$ is uniquely determined by $\mathbf{w}$, $\mathrm{p}(\mathbf{w}|\mathbf{S}) \propto \sqrt{|\mathbf{S}|}\exp(-\frac{1}{2}\mathbf{w}^T\mathbf{S}\mathbf{w})$ captures our Gaussian scale mixture prior, and we assume an independent prior for each $s_j = \frac{1}{\sigma_{w,j}^2}$, $\mathrm{p}(s_j|\epsilon^2) \propto \exp\left(-\frac{s_j}{2\frac{1}{\epsilon^2}}\right)$, as in [1]. In practice, we assume that the wavelet coefficient variance $\sigma_{w,j}^2$ is estimated from the magnitudes of each complex coefficient pair in $\mathbf{w}$:

$$\sigma_{w,2j}^2 = \sigma_{w,2j-1}^2 = \mathbb{E}\left[|w_{2j}|^2\right] = \frac{1}{2}(w_{2j}^2 + w_{2j-1}^2) + \epsilon^2 \tag{8}$$

This gives the following negative log posterior:

$$\begin{aligned} J(\mathbf{w},\mathbf{S}) &= -2\log(\mathrm{p}(\mathbf{w}|\mathbf{y})) \\ &= \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}(\mathbf{y}-\boldsymbol{\lambda})\right\|_2^2 + \mathbf{w}^T\mathbf{S}\mathbf{w} - \log(|\mathbf{S}|) + \epsilon^2\mathrm{tr}(\mathbf{S}) \end{aligned} \tag{9}$$

Next, we allow our estimate of the Gaussian noise variance σ_g^2 to be set using our regularization parameter ν^2. Using $\boldsymbol{\Psi}_\mathbf{y}$ as our estimate of the Poisson variance $\boldsymbol{\lambda}$ in the wavelet domain, we introduce $\boldsymbol{\Sigma}_w$ and add a Majorization Minimization terms in $\mathbf{z}$ to rewrite (9):

$$\begin{aligned} J(\mathbf{w},\mathbf{S},\mathbf{z}) &= \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}(\mathbf{y}_b - \boldsymbol{\Phi}\mathbf{w})\right\|_2^2 \\ &+ \mathbf{w}^T\mathbf{S}\mathbf{w} - \log(|\mathbf{S}|) + \epsilon^2\mathrm{tr}(\mathbf{S}) \\ &+ (\mathbf{w}-\mathbf{z})^T\boldsymbol{\Lambda}_\alpha\boldsymbol{\Sigma}_w^{-1}(\mathbf{w}-\mathbf{z}) - \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\Phi}(\mathbf{w}-\mathbf{z})\right\|_2^2 \end{aligned} \tag{10}$$

Expanding the first and last terms in (10), we have

$$\begin{aligned} J(\mathbf{w},\mathbf{S},\mathbf{z}) &= \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}\mathbf{y}_b\right\|_2^2 - 2\mathbf{y}_b^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathbf{w} + \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\Phi}\mathbf{w}\right\|_2^2 \\ &+ \mathbf{w}^T\mathbf{S}\mathbf{w} - \log(|\mathbf{S}|) + \epsilon^2\mathrm{tr}(\mathbf{S}) \\ &+ (\mathbf{w}-\mathbf{z})^T\boldsymbol{\Lambda}_\alpha\boldsymbol{\Sigma}_w^{-1}(\mathbf{w}-\mathbf{z}) \\ &- \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\Phi}\mathbf{w}\right\|_2^2 + 2\mathbf{z}^T\boldsymbol{\Phi}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi}\mathbf{w} \\ &- \left\|\sqrt{\boldsymbol{\Sigma}^{-1}}\boldsymbol{\Phi}\mathbf{z}\right\|_2^2 \end{aligned} \tag{11}$$

Assuming $\boldsymbol{\Sigma}^{-1}\boldsymbol{\Phi} = \boldsymbol{\Phi}\boldsymbol{\Sigma}_w^{-1}$ and simplifying, we rewrite (11)

$$\begin{aligned} J(\mathbf{w},\mathbf{S},\mathbf{z}) &= -2\mathbf{y}_b^T\boldsymbol{\Phi}\boldsymbol{\Sigma}_w^{-1}\mathbf{w} + \mathbf{w}^T\mathbf{S}\mathbf{w} - \log(|\mathbf{S}|) + \epsilon^2\mathrm{tr}(\mathbf{S}) \\ &+ (\mathbf{w}-\mathbf{z})^T\boldsymbol{\Lambda}_\alpha\boldsymbol{\Sigma}_w^{-1}(\mathbf{w}-\mathbf{z}) \\ &+ 2\mathbf{z}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Sigma}_w^{-1}\mathbf{w} - \mathbf{z}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Sigma}_w^{-1}\mathbf{z} \end{aligned} \tag{12}$$

We minimize eq. (10) by differentiating with respect to $\mathbf{w}, \mathbf{S}$ and $\mathbf{z}$, setting to 0, and multiplying both sides by $\boldsymbol{\Sigma}_w$ where appropriate. This gives rise to the update rules in eqs. (1), (2), and (4).

$$\frac{\partial J(\mathbf{w},\mathbf{S},\mathbf{z})}{\partial \mathbf{w}} = 2[-\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Phi}^T\mathbf{y}_b + \mathbf{S}\mathbf{w} + \boldsymbol{\Sigma}_w^{-1}(\boldsymbol{\Lambda}_\alpha - \boldsymbol{\Phi}^T\boldsymbol{\Phi})\mathbf{z}] \tag{13}$$

$$\frac{\partial J(\mathbf{w},\mathbf{S},\mathbf{z})}{\partial \mathbf{z}} = 2[\boldsymbol{\Sigma}_w^{-1}(\boldsymbol{\Lambda}_\alpha - \boldsymbol{\Phi}^T\boldsymbol{\Phi})(\mathbf{w}-\mathbf{z})] \tag{14}$$

$$\frac{\partial J(\mathbf{w},\mathbf{S},\mathbf{z})}{\partial s_j} = \left(-\frac{1}{s_j} + \sigma_{w,j}^2\right) \tag{15}$$

Last, we've incorporated a spatial domain prior in our algorithm to mitigate the effects of a large-support PSF and the smoothing effects of a wavelet-regularized solution. By thresholding the signal intensities below 0 to 0 at each iteration, we promote much sharper edges at the boundary of the fluorescent object.

IV. Choice of the Parameters

The parameters $\mathbf{b}$ and σ_g^2 are estimated by computing the mean and variance of the K smallest-mean non-overlapping blocks (4x4x4) in a given volume of measurement $\mathbf{y}$, with the final estimates for $\mathbf{b}$ and σ_g^2 the average mean and average variance of these blocks, respectively. We've made a correction to the estimate of $\mathbf{b}$ (+1.5) for all cases, which produced more accurate estimates using when experimenting with our ground truth data. We've tuned K using information about the relative sparsity of the volumes, as some data sets have more zero-blocks than others. a The parameter ν^2 is chosen to decrease geometrically by a factor of 0.64 on each iteration from a value 20% above σ_g^2 to a stopping value just below σ_g^2. This ensures a sparse solution while mitigating the risk of over-regularization. The parameter ϵ^2 is varied in the same way but from a value approximately 10 times the largest observation in y, with a stopping value of σ_g^2. This ensures that the penalty function converges to a sparse solution, while keeping a near quadratic penalty $J(\mathbf{w},\mathbf{z})$ in early iterations so as not to become trapped near undesirable local minima in early iterations.

Convergence is assumed when the relative error $\frac{\|\mathbf{x}_{n+1}-\mathbf{x}_n\|_2}{\|\mathbf{x}_n\|_2} \leq .1\%$, and the volumes look reasonable.

References

[1] Y. Zhang and N. Kingsbury "Fast L0-based Image Deconvolution with Variational Bayesian Inference and Majorization-Minimization".

[2] Y. Zhang and N. Kingsbury "Improved bounds for subband-adaptive iterative shrinkage/thresholding algorithms.", IEEE transactions on image processing, Vol. 22, pp. 1373-1381 (2013).

[3] J. Portilla, S. Vasily, and M. J. Wainwright "Image denoising using scale mixtures of Gaussians in the wavelet domain.", IEEE transactions on image processing, Vol. 12, pp. 1338-1351 (2003).