Fast variational Bayesian signal recovery in the presence of Poisson-Gaussian Noise.

Y. Marnissi[1], Y. Zheng[2], and J.-C. Pesquet[1]

[1] Université Paris-Est, LIGM, UMR CNRS 8049, Champs sur Marne, France
[2] IBM Research, China
Presented by: Amel Benazza-Benyahiya[3]
[3] COSIM Lab., SUP'COM, Carthage Univ., Cité Té chnologique des Communications, Tunisia

March 22[th] 2016

Mixed Poisson Gaussian noise

General model

We observe $\mathbf{y} \in \mathbb{R}^N$ according to the following model

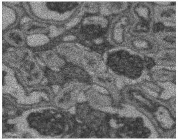
$$\mathbf{y} = \mathbf{z} + \mathbf{w}.$$

- $\mathbf{z} \in \mathbb{R}^N$ ⇝ Poisson noise ($\mathbf{z} \sim \mathcal{P}(\mathbf{Hx})$)
- $\mathbf{x} \in \mathbb{R}^Q$ ⇝ unknown original signal
- $\mathbf{H} \in \mathbb{R}^{N \times Q}$ ⇝ observation operator
- $\mathbf{w} \in \mathbb{R}^N$ ⇝ additive Gaussian noise ($\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2)$)

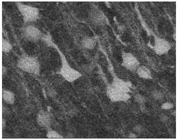
Where?

- CCD camera images [Healey et al. 1994]
- Medical images [Nichols et al. 2002]
- Biological images (fluorescence microscopy) [Pawley 1994]
- Astronomical images [Benvenuto et al. 2008]

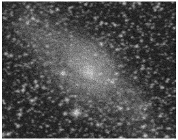
Mixed Poisson Gaussian noise



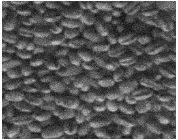
$x_{\max} = 20$
$\sigma^2 = 9$
$\mathbf{H}$ blur with PSF
h: Uniform 5×5



$x_{\max} = 60$
$\sigma^2 = 36$
$\mathbf{H}$ blur with PSF
h: Gaussian 9×9
std 0.5



$x_{\max} = 100$
$\sigma^2 = 36$
$\mathbf{H}$ blur with PSF
h: Uniform 3×3



$x_{\max} = 150$
$\sigma^2 = 40$
$\mathbf{H}$ blur with PSF
h: Gaussian 7×7
std 1

Objective?

⇝ Provide an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ from the collected data $\mathbf{y}$.

Bayesian formulation

OBSERVATION MODEL: $\mathbf{y} = \mathbf{z} + \mathbf{w}$, $\mathbf{z} \sim \mathcal{P}(\mathbf{Hx})$, $\mathbf{w} \sim \mathcal{N}(0, \sigma^2)$

$$\mathrm{p}(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{M} \left(\sum_{n=1}^{+\infty} \frac{e^{-[\mathbf{Hx}]_i} \left([\mathbf{Hx}]_i\right)^n}{n!} \frac{e^{-\frac{1}{2\sigma^2}(y_i - n)^2}}{\sqrt{2\pi\sigma^2}} \right)$$

PRIOR DISTRIBUTIONS:

$$\mathrm{p}(\mathbf{x} \mid \gamma) = \tau \gamma^{\frac{N}{2\kappa}} \exp\left(-\gamma \sum_{j=1}^{Q} \|\mathbf{D}_j \mathbf{x}\|^{2\kappa} \right) \qquad \mathrm{p}(\gamma) \propto \gamma^{\alpha-1} \exp(-\beta\gamma)$$

BAYES FORMULA: $\mathrm{p}(\mathbf{x} \mid \mathbf{y}, \gamma) \propto \mathrm{p}(\mathbf{x} \mid \gamma)\mathrm{p}(\mathbf{y} \mid \mathbf{x})\mathrm{p}(\gamma)$.

✗ DIFFICULTIES: $\mathrm{p}(\mathbf{x} \mid \mathbf{y}, \gamma)$ of a complicated form

- MAP estimation: tuning parameters
- Posterior mean: intractable

✓ SOLUTIONS:

- MCMC methods: computationally expensive
- Variational Bayesian (VB) approaches

Variational Bayesian methodology

Let us denote $\boldsymbol{\Theta}$ the set of the unknown parameters

PRINCIPLE: provide a separable approximation $q(\boldsymbol{\Theta}) = \prod_{j=1}^{J} q(\Theta_j)$ of the true posterior distribution $\mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})$

OPTIMIZATION PROBLEM:

$$q^{opt} = \underset{q}{\operatorname{argmin}}\ \mathcal{KL}(q(\boldsymbol{\Theta}) \| \mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})) \quad \text{s.t.} \quad \text{q is a p.d.f}$$

where

$$\mathcal{KL}(q(\boldsymbol{\Theta}) \| \mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})) = \int q(\boldsymbol{\Theta}) \frac{q(\boldsymbol{\Theta})}{\mathrm{p}(\boldsymbol{\Theta} \mid \mathbf{y})}\, d\boldsymbol{\Theta}$$

An analytical solution (classical variational Bayesian methods) is :

$$(\forall j \in \{1, \ldots, J\}) \;\; q(\Theta_j) \propto \exp\left(\langle \mathrm{p}(\mathbf{y}, \boldsymbol{\Theta}) \rangle_{\prod_{i \neq j} q(\Theta_i)} \right),$$

where $\langle \cdot \rangle_{\prod_{i \neq j} q(\Theta_i)} = \int \cdot \prod_{i \neq j} q(\Theta_i)\, d\Theta_i$

Assumptions and approximations

- SEPARABILITY: We assume that

$$q(\boldsymbol{\Theta}) = q(\mathbf{x})q(\gamma)$$

- GENERALIZED ANSCOMBE TRANSFORM (GAST) APPROXIMATION
  The likelihood of vector $\tilde{\mathbf{y}} \in \mathbb{R}^M$ with components $(\tilde{y}_i)_{1\le i\le M} = 2\left(\sqrt{y_i + \delta}\right)_{1\le i\le M}$, where $\delta = \frac{3}{8} + \sigma^2$, is approximately given by

$$\mathrm{p}(\tilde{\mathbf{y}} \mid \mathbf{x}) = \prod_{i=1}^{M} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\tilde{y}_i - 2\sqrt{[\mathbf{Hx}]_i + \delta}\right)^2\right)$$

✗ DIFFICULTIES: $\mathrm{p}(\mathbf{x}, \gamma \mid \mathbf{y})$ is of a complicated form

✓ SOLUTIONS: Majorizing approximations

Construction of the majorizing approximations

- MAJORANT FUNCTION OF THE NEG-LOG LIKELIHOOD :
  Let $\mathbf{w} = (w_i)_{1 \leq i \leq M} \in [0, +\infty)^M$.
  Since the function $t \longmapsto \sqrt{t+\delta}$ is concave with a Lipschitz continuous gradient, then $T(\tilde{\mathbf{y}}, \mathbf{x}; \mathbf{w}) = \sum_{i=1}^{M} T_i(\tilde{y}_i, [\mathbf{Hx}]_i\,; w_i)$
  where, for every $i \in \{1, \ldots, M\}$,

$$T_i(\tilde{y}_i, [\mathbf{Hx}]_i\,; w_i) = \frac{1}{2}\tilde{y}_i^2 + 2\left([\mathbf{Hx}]_i + \delta\right) - 2\tilde{y}_i\sqrt{w_i+\delta} - \tilde{y}_i\,(w_i+\delta)^{-\frac{1}{2}}\left([\mathbf{Hx}]_i - w_i\right) + \frac{1}{4}\delta^{-\frac{3}{2}}\tilde{y}_i\left([\mathbf{Hx}]_i - w_i\right)^2.$$

  is a majorant of $-\log \mathrm{p}(\tilde{\mathbf{y}} \mid \mathbf{x})$.

Construction of the majorizing approximations

- Majorant function of the neg-log prior of $\mathbf{x}$:
  Let $\boldsymbol{\lambda} = (\lambda_j)_{1 \leq j \leq Q} \in [0, +\infty)^Q$.
  Using the following inequality:

$$(\forall \upsilon \geq 0)(\forall \nu > 0) \quad \upsilon^{\kappa} \leq (1-\kappa)\upsilon^{\kappa} + \kappa \upsilon^{\kappa-1}\nu,$$

then $Q(\mathbf{x}, \gamma; \boldsymbol{\lambda}) = \sum_{j=1}^{Q} Q_j(\mathbf{D}_j\mathbf{x}, \gamma; \lambda_j)$ where for every $j \in \{1, \ldots, Q\}$,

$$(\forall j \in \{1, \ldots, Q\}) \quad Q_j(\mathbf{D}_j\mathbf{x}, \gamma; \lambda_j) = \gamma \frac{\kappa \|\mathbf{D}_j\mathbf{x}\|^2 + (1-\kappa)\lambda_j}{\lambda_j^{1-\kappa}}.$$

is a majorant of $-\log \mathrm{p}(\mathbf{x})$.

Construction of the majorizing approximations

Let

$$L(\boldsymbol{\Theta}|\tilde{\mathbf{y}};\mathbf{w},\boldsymbol{\lambda}) = C(\tilde{\mathbf{y}})\exp\left[-T(\tilde{\mathbf{y}},\mathbf{x};\mathbf{w}) - Q(\mathbf{x},\gamma;\boldsymbol{\lambda})\right]\mathrm{p}(\gamma)$$

where $C(\tilde{\mathbf{y}}) = \mathrm{p}(\tilde{\mathbf{y}})^{-1}(2\pi)^{-M/2}\tau\gamma^{\frac{N}{2\kappa}}$

Then

- $\mathrm{p}(\boldsymbol{\Theta} \mid \tilde{\mathbf{y}}) \geqslant L(\boldsymbol{\Theta}|\tilde{\mathbf{y}};\mathbf{w},\boldsymbol{\lambda})$
- $\mathcal{KL}(q(\boldsymbol{\Theta})\|\mathrm{p}(\boldsymbol{\Theta} \mid \tilde{\mathbf{y}})) \leq \mathcal{KL}(q(\boldsymbol{\Theta})\|L(\boldsymbol{\Theta}|\tilde{\mathbf{y}};\mathbf{w},\boldsymbol{\lambda}))$

To recap...

Let $\mathbf{\Theta}(\mathbf{x}, \gamma)$, $q(\mathbf{\Theta}) = q(\mathbf{x})q(\gamma)$

OPTIMIZATION PROBLEM:

$$q(\mathbf{\Theta}) = \underset{q}{\mathrm{argmin}} \mathcal{KL}(q(\mathbf{\Theta}) \| \mathrm{p}(\mathbf{\Theta} \mid \tilde{\mathbf{y}})) \quad (1)$$

MAJORATION: $\mathcal{KL}(q(\mathbf{\Theta}) \| \mathrm{p}(\mathbf{\Theta} \mid \tilde{\mathbf{y}})) \leq \mathcal{KL}(q(\mathbf{\Theta}) \| L(\mathbf{\Theta} | \tilde{\mathbf{y}}; \mathbf{w}, \boldsymbol{\lambda}))$

Then Thus, Problem (1) can be solved by alternating the following steps:

- Mimimize $\mathcal{KL}(q(\mathbf{\Theta}) \| L(\mathbf{\Theta} | \tilde{\mathbf{y}}; \mathbf{w}, \boldsymbol{\lambda}))$ w.r.t. $q(\mathbf{x})$,
- Update the auxiliary variable $\mathbf{w}$ in order to minimize $\mathcal{KL}(q(\mathbf{\Theta}) \| L(\mathbf{\Theta} | \tilde{\mathbf{y}}; \mathbf{w}, \boldsymbol{\lambda}))$,
- Update the auxiliary variable $\boldsymbol{\lambda}$ in order to minimize $\mathcal{KL}(q(\mathbf{\Theta}) \| L(\mathbf{\Theta} | \tilde{\mathbf{y}}; \mathbf{w}, \boldsymbol{\lambda}))$,
- Mimimize the $\mathcal{KL}(q(\mathbf{\Theta}) \| L(\mathbf{\Theta} | \tilde{\mathbf{y}}; \mathbf{w}, \boldsymbol{\lambda}))$ w.r.t. $q(\gamma)$.

Updating $q(\mathbf{x})$

$$
\begin{aligned}
q^{k+1}(\mathbf{x}) &\propto \exp\left(\left\langle \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^k,\boldsymbol{\lambda}^k)\right\rangle_{q^k(\gamma)}\right) \\
&\propto \exp\left(\int \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^k,\boldsymbol{\lambda}^k) q^k(\gamma) d\gamma\right) \\
&= \mathcal{N}(\mathbf{x};\boldsymbol{m}_{k+1},\boldsymbol{\Sigma}_{k+1})
\end{aligned}
$$

$$
\begin{cases}
\boldsymbol{\Sigma}_{k+1}^{-1} &= \frac{1}{2}\delta^{-\frac{3}{2}}\mathbf{H}^\top \mathrm{Diag}(\tilde{\mathbf{y}})\mathbf{H} + 2\mathrm{E}_{q^k(\gamma)}\mathbf{D}^\top\boldsymbol{\Lambda}^k\mathbf{D}, \\
\boldsymbol{m}_{k+1} &= \boldsymbol{\Sigma}_{k+1}\mathbf{H}^\top\mathbf{u},
\end{cases} \tag{1}
$$

where

- $\mathbf{u} = (u_i)_{1\leqslant i\leqslant M}$, $u_i = \tilde{y}_i(w_i^k+\delta)^{-\frac{1}{2}} + \frac{1}{2}\tilde{y}_i\delta^{-\frac{3}{2}}w_i^k - 2$
- $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are $\left(\kappa(\lambda_j^k)^{\kappa-1}\mathbf{I}_S\right)_{1\leq j\leq Q}$.

Updating $q(\mathbf{x})$

$$
\begin{aligned}
q^{k+1}(\mathbf{x}) &\propto \exp\left(\left\langle \ln L(\mathbf{x}, \gamma, \mathbf{y}; \mathbf{w}^k, \boldsymbol{\lambda}^k)\right\rangle_{q^k(\gamma)}\right) \\
&\propto \exp\left(\int \ln L(\mathbf{x}, \gamma, \mathbf{y}; \mathbf{w}^k, \boldsymbol{\lambda}^k) q^k(\gamma) d\gamma\right) \\
&= \mathcal{N}(\mathbf{x}; \boldsymbol{m}_{k+1}, \boldsymbol{\Sigma}_{k+1})
\end{aligned}
$$

$$
\begin{cases}
\boldsymbol{\Sigma}_{k+1}^{-1} &= \frac{1}{2}\delta^{-\frac{3}{2}}\mathbf{H}^\top \mathrm{Diag}(\tilde{\mathbf{y}})\mathbf{H} + 2\mathrm{E}_{q^k(\gamma)}\mathbf{D}^\top \boldsymbol{\Lambda}^k \mathbf{D}, \\
\boldsymbol{m}_{k+1} &= \boldsymbol{\Sigma}_{k+1}\mathbf{H}^\top \mathbf{u},
\end{cases} \tag{1}
$$

Implementation issues:

- $\boldsymbol{\Sigma}_{k+1}$ is approximated by a diagonal matrix
- $\boldsymbol{m}_{k+1}$ is approximated iteratively using conjugate gradient

Updating $\mathbf{w}$ and $\boldsymbol{\lambda}$

$$\boldsymbol{w}^{k+1} = \underset{\mathbf{w}}{\operatorname{argmin}} \mathcal{KL}(q^{k+1}(\mathbf{x})q^{k}(\gamma) \| L(\boldsymbol{\Theta}|\tilde{\mathbf{y}}; \mathbf{w}^{k+1}, \boldsymbol{\lambda}))$$

$$= \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{M} T_i(\tilde{y}_i, [\mathbf{H}\boldsymbol{m}_{k+1}]_i \,; w_i),$$

$$= \max\{[\mathbf{H}\boldsymbol{m}_{k+1}]_i \,, 0\}$$

$$\lambda_j^{k+1} = \underset{\lambda_j \in [0,+\infty)}{\operatorname{argmin}} \mathcal{KL}(q^{k+1}(\mathbf{x})q^{k}(\gamma) \| L(\boldsymbol{\Theta}|\tilde{\mathbf{y}}; \mathbf{w}^{k+1}, \boldsymbol{\lambda})).$$

$$= \mathrm{E}_{q^{k+1}(\mathbf{x})} \left[\|\mathbf{D}_j \mathbf{x}\|^2 \right]$$

$$= \|\mathbf{D}_j \mathbf{m}_{k+1}\|^2 + \operatorname{trace}\left[\mathbf{D}_j^\top \mathbf{D}_j \boldsymbol{\Sigma}_{k+1} \right]$$

Updating $q(\gamma)$

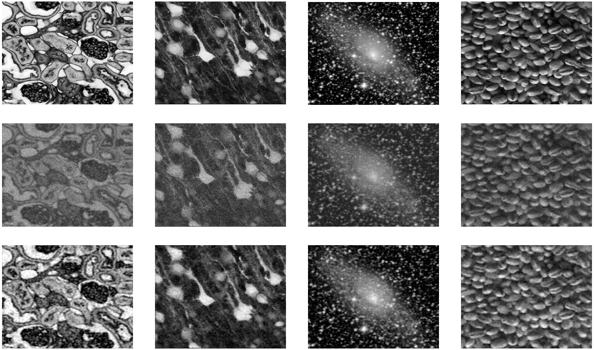
$$
\begin{aligned}
q^{k+1}(\gamma) &\propto \exp\left(\left\langle \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^{k+1},\boldsymbol{\lambda}^{k+1})\right\rangle_{q^{k+1}(\mathbf{x})}\right)\\
&\propto \exp\left(\int \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^{k+1},\boldsymbol{\lambda}^{k+1})q^{k+1}(\mathbf{x})d\mathbf{x}\right)\\
&= \mathcal{G}(\gamma; a_{k+1}, b_{k+1})
\end{aligned}
$$

$$
\begin{cases}
a_{k+1} = \frac{N}{2\kappa} + \alpha = a,\\
b_{k+1} = \sum_{j=1}^{Q}(\lambda_j^{k+1})^{\kappa} + \beta,
\end{cases}
$$

Experiments

- Different count levels $x_{\max}$
- Comparison with variationnal approaches:
  - different PG likelihood approximations: Generalized Anscombe Transform (GAST), the Exponential likelihood (Exp), the Exact likelehood (Exact)
  - different optimization algorithms: Spectral projected gradient [Bajic et al. 1994], Primal-dual splitting algorithm [Chouzenoux et al. 2015]

Visual results



From top to bottom: Original images: $x_{\max} = (20, 60, 100, 150)$. Degraded images: SSIM=(0.232, 0.423, 0.561, 0.586). Restored images with the proposed approach: SSIM=(0.575, 0.653, 0.765, 0.830).

Comparison with variational methods

		MAP (GAST) [PG]	MAP (GAST) [PD]	MAP(EXP) [PD]	MAP (Exact)[PD]	BV (GAST)
First image (350 × 350): $x_{\max} = 20$ h: Uniform 5 × 5, $\sigma^2 = 9$	γ	fixed	fixed	fixed	fixed	automatic
	SNR	13.61	13.60	13.72	13.73	13.80
	Time (s.)	2897	490	3124	48587	29
Second image (257 × 256): $x_{\max} = 60$ h: Gaussian 9 × 9, std 0.5, $\sigma^2 = 36$	γ	fixed	fixed	fixed	fixed	automatic
	SNR	15.35	15.33	15.42	15.43	15.22
	Time (s.)	3168	86	112	612	7
Third image (256 × 256): $x_{\max} = 100$ h: Uniform 3 × 3, $\sigma^2 = 36$	γ	fixed	fixed	fixed	fixed	automatic
	SNR	13.71	13.77	13.81	13.81	14.17
	Time (s.)	2921	578	1060	17027	9
Fourth image (256 × 256): $x_{\max} = 150$ h: Gaussian 7 × 7, std 1, $\sigma^2 = 40$	γ	fixed	fixed	fixed	fixed	automatic
	SNR	20.17	20.11	20.11	20.33	20.43
	Time (s.)	2964	886	3026	43397	14

PG: projected gradient algorithm PD: primal dual algorithm

Simulations performed on an Intel(R) Xeon(R) CPU E5-2630, @ 2.40 GHz, using a Matlab 7 implementation

- Comparable quantitative results with variational approaches
- Regularization parameter automatically tunned
- Competitive computation time

In a nutshell...

- Variational Bayesian method for restoration of images corrupted with mixed Poisson Gaussian noise
- Compared to variational approaches, the proposed method shows:
  - ✓ good quantitative and qualitative results
  - ✓ competitive computation time
- Future work: extension to other PG likelihood approximations and others prior distributions