Compressive sensing for Poisson data

Timo Aspelmeier\textsuperscript{1,2,3}  Uwe Engeland\textsuperscript{3}  Gernot Ebel\textsuperscript{3}

\textsuperscript{1}Institut für Mathematische Stochastik, Georg-August-Universität Göttingen
\textsuperscript{2}Max-Planck-Institut für biophysikalische Chemie, Göttingen
\textsuperscript{3}Scivis wissenschaftliche Bildverarbeitung GmbH, Göttingen

Göttingen, 13 November 2013
Compressive sensing for Poisson data

Introduction

Usual compressive sensing setup:

\[ y = Af + \epsilon \]

- \( y \in \mathbb{R}^m \) \hspace{1cm} measurements
- \( f \in \mathbb{R}^n \) \hspace{1cm} signal of interest
- \( A \in \mathbb{R}^{m \times n} \) \hspace{1cm} measurement matrix
- \( \epsilon \in \mathbb{R}^m \) \hspace{1cm} i.i.d. Gaussian noise, mean 0, variance \( \sigma^2 \)
- \( m \ll n \)
Compressive sensing for Poisson data

Why consider non-Gaussian noise?

In many applications: noise is not Gaussian

Astrophysical image  Fluorescence microscopy image  SPECT image

Poisson noise comes about from discrete quantum nature of light
Compressive sensing for Poisson data is important for

- getting the most out of low-light situations
- dose reduction in medical applications:
  - SPECT
  - PET
  - CT
  - planar X-ray
- faster acquisition times for improved temporal resolution
- ...
Problems with Poisson data

Differences to the Gaussian case

Compressive sensing is not immediately applicable to Poisson data:

- Poisson noise variance
- Constraints on signal $f$
- Constraints on measurement matrix $A$
Problems with Poisson data

Poisson noise variance

Poisson variable $N$:

$$\mathbb{E}N = \lambda$$
$$\text{Var}(N) = \lambda$$

- Non-constant signal dependent variance
- non-concentrated
- cf. Gaussian variable $\epsilon$: $\text{Var}(\epsilon) = \sigma^2 = \text{const.}$

Constancy of variance necessary, e.g. in Haupt, Nowak (2006)
Problems with Poisson data

Physical considerations

No “negative” photons possible:

\[ f_i \geq 0, \quad i = 1, \ldots, n \]

No “photon subtraction”:

\[ A_{ij} \geq 0, \quad i = 1, \ldots, m, j = 1, \ldots, n \]

Flux preservation:

\[ \sum_i (Af)_i \leq C \sum_j f_j, \quad \text{with some } C > 0 \text{ for all feasible } f \]

M. Raginsky et al. (2010)
Problem formulation

Compressibility properties

Problem to solve:

\[ y_i \sim \text{Poisson}((Af)_i + b_i) \]
\[ b_i \geq 0 \quad \text{known background signal} \]

Assume \( f \) is \( p \)-compressible for some orthogonal matrix \( T \):

\[ \theta_k = (Tf)_k \]
\[ |\theta_k| \leq \text{const.} \times k^{-1/p} \quad \text{(ordered)} \]
\[ p < 2 \]
Problem formulation

Sparsity enforced maximum likelihood estimation

Sparsity penalized maximum likelihood estimator

\[ \hat{f} = \arg\min_f L(f; y) + \lambda_0 \| T f \|_0 \]

\[ L(f; y) = \sum_{i=1}^m ((Af)_i + b_i - y_i \log((Af)_i + b_i)) \]

Digression: compressive sensing with benign noise

Unifying strategy

\[ \hat{f} = \text{argmin}_\phi \lambda_1 \| A\phi - y \|_2^2 + \lambda_0 \| T\phi \|_0 \]

Loss function

\[ R(\hat{f}, f) = \frac{1}{n} \| \hat{f} - f \|_2^2 \]

<table>
<thead>
<tr>
<th>Type</th>
<th>( \lambda_0 )</th>
<th>( \lambda_1 )</th>
<th>Error bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>No noise, exact sparsity</td>
<td>1</td>
<td>( \infty )</td>
<td>0 if ( \frac{\log n}{m} \leq \text{const.} \times m )</td>
</tr>
<tr>
<td>No noise, compressibility ( p )</td>
<td>1</td>
<td>( \infty )</td>
<td>( \left( \frac{\log n}{m} \right)^{\frac{1}{p} - \frac{1}{2}} )</td>
</tr>
<tr>
<td>Benign noise, compressibility ( p )</td>
<td>\text{const.}</td>
<td>1</td>
<td>( \left( \frac{\log n}{m} \right)^{\frac{1}{2} - \frac{p}{4}} )</td>
</tr>
</tbody>
</table>
Performance bounds for Poisson data

Random matrix approach

Construct random matrix $A$ by

$$A_{ij} = \begin{cases} 0 & \text{with probability } p \\ \frac{1}{m} & \text{with probability } 1 - p \end{cases}$$

- $A$ fulfills all requirements
- $A$ is Rademacher matrix + constant offset
- Rademacher component $\Rightarrow$ RIP
Performance bounds for Poisson data

Random matrix approach

Theorem (Raginsky et al. (2010))

Under appropriate conditions

$$\mathbb{E} \frac{R(\hat{f}, f)}{\left(\sum_{j=1}^{n} f_j\right)^2} \leq O(m) \min_{1 \leq k \leq k^*} \left[ k^{-2\alpha} + \frac{k}{n} + \frac{k \log_2 n}{\sum_j f_j} \right] + O \left( \frac{\log(n/m)}{m} \right)$$

with high probability, where

$$k^* = \text{const.} \times \frac{m}{\log_2 n},$$

$$\alpha = \frac{1}{p} - \frac{1}{2}. $$

Timo Aspelmeier, Uwe Engeland, Gernot Ebel

Compressive sensing for Poisson data
Performance bounds for Poisson data
Random matrix approach: consequences

Surprising consequences:

- Low intensities $\sum_j f_j$: performance bound is

$$\sim \mathcal{O}(m) \left( \frac{\log n}{\sum_j f_j} \right)^{2\alpha/(2\alpha+1)}$$

- Error bound is always $\mathcal{O}(m)$, not $\mathcal{O}((\log m/n)^\gamma)$ with some $\gamma > 0$
Performance bounds for Poisson data

Deterministic matrices

In practice often (see description of SPECT below):

- Measurement matrix $A$ is not random
- $A$ is given by physical requirements
- $A$ can only be “designed” within certain limits

Even for random matrices:

- Performance bounds appear of limited use
- Very low light situations seem particularly dire
Performance bounds for Poisson data

Experimental evidence

Fourier-sparse image

same image with Poisson noise

reconstruction from 100 random points

reconstruction from 400 random points
Performance bounds for Poisson data

Weaker performance bounds

Alternative approach:

- take $A$ as given
- use random selection matrix approach
- use weaker measure of error

Random selection: e.g., E. Candès (2006)
Performance bounds for Poisson data

Weaker performance bounds

With random selection matrix $S$:

$$y_i \sim \text{Poisson}((SAf)_i + b_i)$$

$A \in \mathbb{R}^{n' \times n}$  

$m \ll n$  

$S \in \mathbb{R}^{m \times n'}$,  

$n' = \mathcal{O}(n)$

SA fulfills requirements for Poisson measurement matrix

Weaker loss function: Kullback-Leibler divergence

$$R^{KL}(\hat{f}, f) = \sum_i \left( (A\hat{f})_i - (Af)_i - (Af)_i \log \frac{(A\hat{f})_i}{(Af)_i} \right)$$
Theorem (Aspelmeier et al. (2011, 2013))

Under appropriate conditions,

\[
\mathbb{E} R^{KL}(\hat{f}, f) \leq \text{const.} \times \left( \frac{(J\sqrt{n})^{2/(2\alpha-1)} \log n}{m} \right)^{\frac{2\alpha-1}{2\alpha+1}},
\]

with high probability, where

\[
J = \max_{i,j} |(AT^{-1})_{ij}|
\]

(incoherence).
Performance bounds for Poisson data

CT

For CT:

\[ y_i \sim \text{Poisson} \left( \sum_j B_{ij} \exp \left( -\alpha (A f)_j \right) + b_i \right) \]

\[ \alpha > 0 \]

\[ B_{ij} : \text{ Point spread function of detector} \]

Then

\[ \mathbb{E} R_{KL}(\hat{f}, f) \leq \text{const.} \times \left( \frac{(J \sqrt{n})^{2/(2\alpha - 1)} \log n}{m} \right)^{\frac{2\alpha - 1}{2\alpha + 1}}, \]

still holds (with different constants)
Performance bounds for Poisson data

Selection matrix approach: consequences

Consequences:

- No special low-intensity results
- RIP not feasible; result contains incoherence explicitly
- Error bound is $O((J\sqrt{n})^{2/(2\alpha+1)}(\log n/m)^{\frac{2\alpha-1}{2\alpha+1}})$
- Non-optimal incoherence $\Rightarrow$ power law gain
- $\frac{2\alpha-1}{2\alpha+1} > 0 \Rightarrow p < 1$: strong compressibility needed
- Error bound exponent is worse than for Gaussian compressive sensing $(2\alpha/(2\alpha + 1))$
- $R^{KL}(\hat{f}, f)$ measures error in projections
- Additional regularization may be necessary
Performance bounds for Poisson data

Comparison of exponents

![Graph showing error bound exponents for compressibility](image)
Algorithmic challenges

Nonsmooth convex optimization

Original problem:

\[ \hat{f} = \arg\min_{f'} L(f'; y) + \lambda_0 \| T f' \|_0 \]

Nonconvex, therefore \( \ell_0 \to \ell_1 \):

\[ \hat{f} = \arg\min_{f'} L(f'; y) + \lambda_1 \| T f' \|_1 \]

More difficult to solve than \( \ell_2-\ell_1 \) problems:

\[ \hat{f} = \arg\min_{f'} \| f' - y \|_2^2 + \lambda_1 \| T f' \|_1 \]
Algorithmic challenges
Nonsmooth convex optimization

Still difficult to solve numerically since
- $\|\cdots\|_1$ is nonsmooth
- $\nabla L$ is not Lipshitz-continuous

Some possibilities (not complete):
- Gradient-based EM: problematic
- Beck-Teboulle A. Beck and M. Teboulle (2009)
- Chambolle-Pock A. Chambolle and Th. Pock (2011)
- SPIRAL-TAP Z. Harmany et al. (2009, 2011)
- SpaRSA S. Wright et al. (2009)
- Anscombe transform methods C. Chaux et al. (2007), F.-X. Dupé et al. (2009)
Applications

SPECT: principles

SPECT measurement matrix $A$:

- $\approx$ discretized Radon transform
- conical view field (collimator) $\leftrightarrow$ distance-dependent PSF, truly 3d

Image acquisition:

- very low counts ($\approx 100$ per pixel)
- long acquisition times
SPECT reconstructions of a Fourier phantom

Phantom and projections

Numerical phantom with sparse Fourier coefficients

1 of 35 GATE simulated SPECT projections
SPECT reconstructions of a Fourier phantom

Reconstructions

Penalized likelihood
$$\text{argmin}_{f'} L(f'; y) + \lambda_1 \Phi(f)$$

Original
$$\text{argmin}_{f'} L(f'; y) + \lambda_2 \| Tf' \|_1$$

Compressive sensing

Timo Aspelmeier, Uwe Engeland, Gernot Ebel
Conclusion

Summary:

- Compressive Sensing for Poissonian data problems:
  - Non-concentrated, data dependent noise
  - Physical constraints on $A$ and $f$
  - In practice: given measurement matrix $A$
  - Diverging error bounds for $\ell_2$ loss
- Alternative approach
  - Random measurement selection
  - Kullback-Leibler divergence as loss
  - Weaker performance bounds

Conclusion:

- CS works (less effectively) for Poissonian data
- Potential applications in medical imaging and many other fields
- Future work: conventional regularization as “add on”
Conclusion

Summary:

- Compressive Sensing for Poissonian data problems:
  - Non-concentrated, data dependent noise
  - Physical constraints on $A$ and $f$
  - In practice: given measurement matrix $A$
  - Diverging error bounds for $\ell_2$ loss
- Alternative approach
  - Random measurement selection
  - Kullback-Leibler divergence as loss
  - Weaker performance bounds

Conclusion:

- CS works (less effectively) for Poissonian data
- Potential applications in medical imaging and many other fields
- Future work: conventional regularization as “add on”
Planar X-ray dose reduction
Image noise at different doses

25%  32%

50%  100%

Timo Aspelmeier, Uwe Engeland, Gernot Ebel
Compressive sensing for Poisson data
Planar X-ray dose reduction
Image noise at different doses
Planar X-ray dose reduction
Noise removal

50% processed   100%   50% original
Planar X-ray dose reduction

Noise removal

25% processed  100%  25% original

Aspelmeier, Ebel, Engeland 2012