Statistical regularization theory for Inverse Problems with Poisson data

Frank Werner\textsuperscript{1,2}, joint with Thorsten Hohage

\textsuperscript{1}Statistical Inverse Problems in Biophysics Group
Max Planck Institute for Biophysical Chemistry

\textsuperscript{2}Felix Bernstein Institute for Mathematical Statistics in the Biosciences
University of Göttingen
Outline

1 Introduction

2 A continuous model

3 Regularization methods

4 Examples for regularization methods

5 Conclusion
Outline

1 Introduction
2 A continuous model
3 Regularization methods
4 Examples for regularization methods
5 Conclusion
Why focusing on Poisson data?

- In various applications measurements are photon counts:
  - Fluorescence microscopy
  - Astronomical imaging
  - X-ray diffraction imaging
  - Positron Emission Tomography
  - ...

- At low energies the quantization of energy is the main source of noise
- Given an ideal photon detector, the data is purely Poisson distributed
Discrete model

- Suppose the imaging procedure is modelled by a mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$
- Let $u^\dagger \in \mathbb{R}^n$ denote the exact solution we seek for and $g^\dagger := F(u^\dagger)$, require $g^\dagger \geq 0$
- For the data $g_{\text{obs}} \in \mathbb{R}^m$ the value $g_{i,\text{obs}}$ is the number of photon counts in detector region $i \in \{1, \ldots, m\}$
- In the ideal case $g_{\text{obs}} \in \mathbb{R}^m$ is a random variable such that $g_{i,\text{obs}} \sim \text{Poi} \left(g_i^\dagger \right)$, e.g.

\[
P \left[ g_{i,\text{obs}} = k \right] = \frac{(g_i^\dagger)^k}{k!} \exp \left(-g_i^\dagger \right).
\]
Outline

1 Introduction

2 A continuous model

3 Regularization methods

4 Examples for regularization methods

5 Conclusion
Continuous model I

- Now $F : \mathcal{X} \to \mathcal{Y}$ with Banach spaces $\mathcal{X}$ and $\mathcal{Y} \subset L^1(M)$
- Consequently $u^\dagger \in \mathcal{X}$ and $g^\dagger := F(u^\dagger) \in L^1(M)$, require $g^\dagger \geq 0$

Deterministic approach

Suppose the observed data $g^{\text{obs}} \in \mathcal{Y}$ satisfies a noise bound

$$\left\| g^{\text{obs}} - g^\dagger \right\|_{\mathcal{Y}} \leq \delta.$$ 

Alternatively, the norm $\| \cdot - \cdot \|_{\mathcal{Y}}$ could be replaced by a different norm or a general loss $d$. 
Continuous model II

- In the ideal case, the data still consists of photon counts.
- Say the total number of observed photons is \( n \) and their positions are \( x_i \in \mathbb{M} \).
- \( n \) can be influenced by the 'exposure time', mathematically described by a scaling factor \( t > 0 \).
- Associate the measure \( \tilde{G}_t = \sum_{i=1}^{n} \delta_{x_i} \).

Statistic approach

The observed data is a scaled Poisson process \( G_t = \tilde{G}_t / t \) with intensity \( g^\dagger \), i.e. the measure \( \tilde{G}_t \) satisfies the following axioms:

1. For each choice of disjoint, measurable sets \( A_1, ..., A_n \subset \mathbb{M} \) the random variables \( \tilde{G}_t (A_j) \) are stochastically independent.

2. \( \mathbb{E} \left[ \tilde{G}_t (A) \right] = \int_A t g^\dagger \, dx \) for all \( A \subset \mathbb{M} \) measurable.
A continuous model

Deterministic vs. statistic model

- Deterministic model:
  - Clear definition of the noise level, but ...
  - ... the relation to a Poisson distribution is lost!

- Statistic model:
  - Poisson distribution incorporated, in fact it holds

\[ \tilde{G}_t(A) \sim \text{Poi} \left( t \int_A g^\dagger \, dx \right) \]

for all measurable \( A \subset M \), but ...

- ... \( G_t \notin \mathcal{Y} \) and no clear definition of the noise level so far!

- Note: Similar statistic model is used by Cavalier & Koo 2002, Antoniadis & Bigot 2006.
A continuous model

Statistic model: noise level 1

- For a function $g$ let

$$\int_{\mathcal{M}} g \, d\tilde{G}_t := \sum_{i=1}^{n} g(x_i)$$

- Then

$$E \left[ \int_{\mathcal{M}} g \, dG_t \right] = \int_{\mathcal{M}} gg^\dagger \, dx,$$

$$\text{Var} \left[ \int_{\mathcal{M}} g \, dG_t \right] = \frac{1}{t^2} E \left[ \int_{\mathcal{M}} g^2 \, d\tilde{G}_t \right] = \frac{1}{t} \int_{\mathcal{M}} g^2 g^\dagger \, dx.$$  

- Thus any bounded linear functional of $g^\dagger$ can be estimated unbiasedly with a variance proportional to $\frac{1}{t}$.

- This suggests that the noise level should be proportional to $1/\sqrt{t}$. 

Frank Werner, Göttingen, Germany
Inverse Problems with Poisson data
May 26th, IPMS 2014 10 / 27
Any bounded linear functional of $g^{\dagger}$ can be estimated unbiasedly with a variance proportional to $\frac{1}{t}$.

For our analysis, such a property is needed uniformly!

The following result is based on the work of Renaud-Bouret 2003.

**Uniform concentration inequality (W., Hohage 2012)**

Suppose $\mathbb{M} \subset \mathbb{R}^d$ is bounded & Lipschitz, $s > d/2$ and set

$$\mathcal{G}(R) := \{ g \in H^s(\mathbb{M}) : \| g \|_{H^s} \leq R \}.$$

Then $\exists \ c = c (\mathbb{M}, s, \| g^{\dagger} \|_{L^1}) > 0$ such that

$$P \left[ \sup_{g \in \mathcal{G}(R)} \left| \int_{\mathbb{M}} g \left( dG_t - g^{\dagger} \, dx \right) \right| \geq \frac{\rho}{\sqrt{t}} \right] \leq \exp \left( -\frac{\rho}{cR} \right)$$

for all $R \geq 1$, $t \geq 1$ and $\rho \geq cR$. 
Outline

1. Introduction

2. A continuous model

3. Regularization methods

4. Examples for regularization methods

5. Conclusion
Some notation

- Replace the discontinuous mapping $F^{-1}$ by **continuous** approximations $R_\alpha$
- Often solutions restricted to $\mathcal{B} \subset \mathcal{X}$
  - Deterministic $\rightsquigarrow R_\alpha : \mathcal{Y} \to \mathcal{B}$
  - Statistic $\rightsquigarrow R_\alpha : \mathcal{M} (\mathcal{M}) \to \mathcal{B}$ where $\mathcal{M} (\mathcal{M}) \triangleq$ space of all measures
- $\alpha$ chosen by a parameter choice rule $\bar{\alpha}$
  - Deterministic $\rightsquigarrow \bar{\alpha} : (0, \infty) \times \mathcal{Y} \to (0, \infty)$
  - Statistic $\rightsquigarrow \bar{\alpha} : (0, \infty) \times \mathcal{M} (\mathcal{M}) \to (0, \infty)$
- We aim for ’convergence’ w.r.t. a **loss** $d : \mathcal{B} \times \mathcal{B} \to [0, \infty)$ with $d(u, u) = 0$ for all $u \in \mathcal{B}$
- Typical examples: Bregman distance

$$d \left( u, u^\dagger \right) = D_{R_{\bar{\alpha}}}^{u^*} \left( u, u^\dagger \right) := R(u) - R \left( u^\dagger \right) - \left\langle u^*, u - u^\dagger \right\rangle$$

where $u^* \in \partial R \left( u^\dagger \right) \subset \mathcal{X}'$ or norm $d \left( u, u^\dagger \right) = \| u - u^\dagger \|_{\mathcal{X}}$. 

Frank Werner, Göttingen, Germany

Inverse Problems with Poisson data

May 26th, IPMS 2014
### Regularization schemes

#### Deterministic Regularization Scheme

\((R_\alpha, \bar{\alpha})\) is called a **deterministic regularization scheme** w.r.t. \(d\) if

\[
\lim_{\delta \to 0} \sup \left\{ d \left( R_{\bar{\alpha}}(\delta, g^{\text{obs}}) \left( g^{\text{obs}} \right), u^\dagger \right) \mid g^{\text{obs}} \in \mathcal{Y}, \| g^{\text{obs}} - F(u^\dagger) \| \leq \delta \right\} = 0
\]

#### Statistical Regularization Scheme

\((R_\alpha, \bar{\alpha})\) is called a (consistent) **statistical regularization scheme** under Poisson data w.r.t. \(d\) if

\[
\forall \varepsilon > 0 : \lim_{t \to \infty} \mathbb{P} \left[ d \left( R_{\bar{\alpha}}(t, G_t) \left( G_t \right), u^\dagger \right) > \varepsilon \right] = 0,
\]

where \(G_t\) is a scaled Poisson process with intensity \(F(u^\dagger)\).
Convergence rates

- let $\psi : [0, \infty) \rightarrow [0, \infty), \psi \nearrow, \psi(0) = 0, M \subset \mathcal{B}$.

**Deterministic convergence rates**

$(R_{\alpha}, \bar{\alpha})$ obeys the **deterministic convergence rate** $\psi$ on $M$ w.r.t. $d$ if

$$d \left( R_{\bar{\alpha}}(\delta, g_{\text{obs}}) \left( g_{\text{obs}} \right), u^\dagger \right) = O(\psi(\delta)), \quad \delta \searrow 0$$

for all $u^\dagger \in M$ and $\|g_{\text{obs}} - F(u^\dagger)\|_Y \leq \delta$.

**Statistical convergence rates**

$(R_{\alpha}, \bar{\alpha})$ obeys the **statistical convergence rate** $\psi$ on $M$ w.r.t. $d$ if

$$E \left[ d \left( R_{\bar{\alpha}}(\delta, G_t) \left( G_t \right), u^\dagger \right) \right] = O(\psi(t)), \quad t \rightarrow \infty$$

for all $u^\dagger \in M$ where $G_t$ is a scaled Poisson process with intensity $F(u^\dagger)$. 
Examples for regularization methods

Outline

1 Introduction

2 A continuous model

3 Regularization methods

4 Examples for regularization methods

5 Conclusion
Regularization by projection

- Suppose $F = T$ is bounded, linear and positive definite, for simplicity $\mathcal{X} = \mathcal{Y} = L^2(M)$.
- Regularization by projection: $V_n \subset \mathcal{X}$ with $\dim(V_n) < \infty$ and
  \[
  u_{n}^{\text{proj}} := \underset{u \in V_n}{\text{argmin}} \left\| Tu - g^{\text{obs}} \right\|_Y^2
  \]  
  (1)
- If $\{v_1, ..., v_n\}$ is an orthonormal basis of $V_n$, then
  \[
  u_{n}^{\text{proj}} \in V_n : \quad \langle Tu_{n}^{\text{proj}}, v_j \rangle = \int_M v_j g^{\text{obs}} \, dx, \quad 1 \leq j \leq n.
  \]
- Define $u_{n}^{\text{proj}}$ also in the statistical case by replacing $g^{\text{obs}}$ by $G_t$.
- In principle the norm in (1) can be replaced by any other loss, but ...
- ... for the natural Poissonian choice this leads to problems proving existence of $u_{n}^{\text{proj}}$ and stability.
Examples for regularization methods

Projection methods

Some (simplified) results

\[ u_n^{\text{proj}} \in V_n : \quad \langle Tu_n^{\text{proj}}, v_j \rangle = \int_{\mathbb{M}} v_j \, dG_t, \quad 1 \leq j \leq n. \]

- Cavalier & Koo 2002:
  - \( V_n = \) suitable wavelet space, \( T = \) Radon transform
  - The projection estimator exists and depends continuously on the data
  - Explicit convergence rate as \( t \to \infty \)
- Problem: \( u_n^{\text{proj}} \) in general not non-negative
- Antoniadis & Bigot 2006:
  - \( V_n = \exp(U_n) \) with a suitable Wavelet space \( U_n \)
  - Corresponding estimator (if existent) is always non-negative and depends continuously on the data
  - As \( t \to \infty \), the estimator exists with probability 1.
  - Explicit convergence rate as \( t \to \infty \)
  - Consistency of the estimator is unknown, i.e. it is unclear if \( u_n^{\text{proj}} \) yields a statistical regularization scheme.
Variational regularization I

- Disadvantage of the aforementioned methods: design does not rely on Poisson distribution!
- Different approach: likelihood methods!
  Minimize

\[ u \mapsto S(G_t; F(u)) := - \ln \left( P \left[ G_t \mid \text{the exact density is } F(u) \right] \right) \]

over all admissible \( u \).

- Still ill-posed due to ill-posedness of the original problem. This gives rise to the following variant of Tikhonov regularization:

\[ u_\alpha \in \arg\min_{u \in \mathcal{B}} [S(G_t; F(u)) + \alpha \mathcal{R}(u)] \]

where \( \mathcal{R} \) is a convex penalty term and \( \alpha > 0 \) a regularization parameter.
Variational regularization II

\[
u_{\alpha} \in \arg\min_{u \in B} \left[ S(G_t; F(u)) + \alpha R(u) \right]
\]

- Main issue in the analysis: data fidelity term lacks of a triangle-type inequality!
- References for deterministic regularization properties:
- References for deterministic convergence rates:
  Benning & Burger 2011, Flemming 2010 & 2011 ...
- Here: statistic case.
Data fidelity terms

- Negative log-likelihood for a scaled Poisson process:

\[ S_0 (G_t; g) = \int_M g \, dx - \int_M \ln (g) \, dG_t, \quad g \geq 0 \text{ a.e.} \]

- Ideal data misfit functional for exact data \( g^\dagger \) given by

\[
E [S_0 (G_t; g)] - E [S_0 (G_t; g^\dagger)] = \int_M \left[ g - g^\dagger - g^\dagger \ln \left( \frac{g}{g^\dagger} \right) \right] \, dx
\]

which is the Kullback-Leibler divergence \( KL (g^\dagger; g) \).

- We introduce a shift \( \sigma > 0 \) and consider

\[ S_\sigma (G_t; g) := \int_M g \, dx - \int_M \ln (g + \sigma) \, (dG_t + \sigma \, dx) \]

\[ T (g^\dagger; g) := KL (g^\dagger + \sigma; g + \sigma) \]
Assumptions

### Assumptions on the problem

- \((\mathcal{X}, \tau_{\mathcal{X}})\) top. vector space, \(\tau_{\mathcal{X}}\) weaker than norm topology, and \(\mathcal{B} \subset \mathcal{X}\) closed and convex.
- \(F: \mathcal{B} \rightarrow L^1(\mathbb{M})\) with \(\mathbb{M} \subset \mathbb{R}^d\) bounded & Lipschitz and
  1. \(F: \mathcal{B} \rightarrow L^1(\mathbb{M})\) is \(\tau_{\mathcal{X}} - \tau_\omega\)-sequentially continuous.
  2. \(F(u) \geq 0\) a.e. for all \(u \in \mathcal{B}\).
  3. There exists \(s > d/2\) such that \(F(\mathcal{B})\) is a bounded subset of \(H^s(\mathbb{M})\).

### Assumptions on the method

- \(R: \mathcal{B} \rightarrow (-\infty, \infty]\) is convex, proper and \(\tau_{\mathcal{X}}\)-sequentially lower semicontinuous.
- \(R\)-sublevelsets \(\{u \in \mathcal{X} \mid R(u) \leq M\}\) are \(\tau_{\mathcal{X}}\)-sequentially pre-compact.
Statistical regularization properties

\[ u_\alpha \in \arg\min_{u \in \mathcal{B}} [S_\sigma (G_t; F(u)) + \alpha R(u)] \]

Under those assumptions, a minimizer \( u_\alpha \) exists with probability one.

Regularization properties (Hohage, W. 2014)

\( R_\alpha G_t := u_\alpha \) with any minimizer \( u_\alpha \) equipped with any parameter choice rule \( \bar{\alpha} \) fulfilling

\[
\lim_{t \to \infty} \bar{\alpha}(t, G_t) = 0, \quad \lim_{t \to \infty} \frac{\ln(t)}{\sqrt{t\bar{\alpha}(t, G_t)}} = 0
\]

defines a statistical regularization scheme under Poisson data w.r.t. the Bregman distance.

T. Hohage and F. Werner.  
Inverse Problems with Poisson Data: statistical regularization theory, applications and algorithms. 
In preparation, 2014
**Source condition**

- As the problem is ill-posed, convergence rates can only be obtained for a strict subset $M \subset \mathcal{X}$
- Here the set $M$ is described by a **variational inequality** as source condition:

$$\beta D_{\mathcal{R}}^{u^*} (u, u^\dagger) \leq \mathcal{R}(u) - \mathcal{R}(u^\dagger) + \varphi \left( \mathcal{T} \left( g^\dagger; F(u) \right) \right)$$

(2)

for all $u \in \mathcal{B}$ with $\beta > 0$. $\varphi$ is assumed to fulfill

- $\varphi(0) = 0$,
- $\varphi \uparrow$,
- $\varphi$ concave.

- Now the source set $M = M_{\mathcal{R}}^{\varphi}(\beta)$ consists of all $u^\dagger \in \mathcal{B}$ satisfying (2).
Examples for regularization methods  Statistical convergence rates

Statistical convergence rates

A priori convergence rates (W., Hohage 2012)

Then for $\alpha = \alpha(t)$ chosen appropriately we obtain the statistical convergence rate $\psi(t) = \varphi\left(\frac{1}{\sqrt{t}}\right)$ on $M^\varphi_{\mathcal{R}}(\beta)$ w.r.t. $\mathcal{D}_{\mathcal{R}} u^* (\cdot, u^\dagger)$, i.e.

$$
\mathbf{E} \left[ \mathcal{D}_{\mathcal{R}} u^* \left( u_\alpha, u^\dagger \right) \right] = \mathcal{O} \left( \varphi \left( \frac{1}{\sqrt{t}} \right) \right), \quad t \to \infty.
$$

Under suitable assumptions $\alpha$ can be chosen according to a Lepskiǐ-type balancing principle yielding the same rate up to a log-factor.

F. Werner and T. Hohage.  
Convergence rates in expectation for Tikhonov-type regularization of Inverse Problems with Poisson data.  
*Inverse Problems* 28, 104004, 2012
Outline

1. Introduction
2. A continuous model
3. Regularization methods
4. Examples for regularization methods
5. Conclusion
Presented results

- Regularization theory for inverse problems with Poisson data:
  - Sound mathematical model
  - Definition of regularization properties
  - Convergence rates

- Projection-type estimators

- Tikhonov regularization obeys all those properties under reasonable assumptions

Thank you for your attention!