Statistical regularization theory for Inverse Problems with Poisson data

Frank Werner[1,2], joint with Thorsten Hohage

[1]Statistical Inverse Problems in Biophysics Group
Max Planck Institute for Biophysical Chemistry

[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

Why focusing on Poisson data?

- In various applications measurements are photon counts:
  - Flourescence microscopy
  - Astonomical 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 $F : \mathbb{R}^n \to \mathbb{R}^m$
- Let $u^\dagger \in \mathbb{R}^n$ denote the exact solution we seek for and $g^\dagger := F\left(u^\dagger\right)$, require $g^\dagger \geq 0$
- For the data $g^{\mathrm{obs}} \in \mathbb{R}^m$ the value g_i^{obs} is the number of photon counts in detector region $i \in \{1, ..., m\}$
- In the ideal case $g^{\mathrm{obs}} \in \mathbb{R}^m$ is a random variable such that $g_i^{\mathrm{obs}} \sim \mathrm{Poi}\left(g_i^\dagger\right)$, e.g.

$$\mathbf{P}\left[g_i^{\mathrm{obs}} = k\right] = \frac{\left(g_i^\dagger\right)^k}{k!} \exp\left(-g_i^\dagger\right).$$

Outline

Continuous model I

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

Deterministic approach

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

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

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. $\mathbf{E}\left[\tilde{G}_t(A)\right] = \int_A t g^\dagger \,\mathrm{d}x$ for all $A \subset \mathbb{M}$ measurable.

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

$$\tilde{G}_t(A) \sim \mathrm{Poi}\left(t \int\limits_A g^\dagger \,\mathrm{d}x\right)$$

    for all measurable $A \subset \mathbb{M}$, 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.

Statistic model: noise level I

- For a function g let

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

- Then

$$\mathbf{E}\left[\int_{\mathbb{M}} g \, \mathrm{d}G_t \right] = \int_{\mathbb{M}} g g^\dagger \, \mathrm{d}x,$$

$$\mathbf{Var}\left[\int_{\mathbb{M}} g \, \mathrm{d}G_t \right] = \frac{1}{t^2} \mathbf{E}\left[\int_{\mathbb{M}} g^2 \, \mathrm{d}\tilde{G}_t \right] = \frac{1}{t} \int_{\mathbb{M}} g^2 g^\dagger \, \mathrm{d}x.$$

- 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 $1/\sqrt{t}$.

Statistic model: noise level II

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

$\rightsquigarrow$ 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

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

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

$$\mathbf{P}\left[\sup_{\mathfrak{g}\in\mathfrak{G}(R)} \left|\int_{\mathbb{M}} \mathfrak{g}\left(\mathrm{d}G_t - g^\dagger\,\mathrm{d}x\right)\right| \geq \frac{\rho}{\sqrt{t}}\right] \leq \exp\left(-\frac{\rho}{cR}\right)$$

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

Outline

Some notation

- Replace the discontinuous mapping F^{-1} by **continuous** approximations R_α
- Often solutions restricted to $\mathfrak{B} \subset \mathcal{X}$

◇ Deterministic $\rightsquigarrow$ $R_\alpha : \mathcal{Y} \to \mathfrak{B}$
Statistic $\rightsquigarrow$ $R_\alpha : \mathfrak{M}(\mathbb{M}) \to \mathfrak{B}$ where $\mathfrak{M}(\mathbb{M}) \hat{=}$space of all measures

- α chosen by a parameter choice rule $\bar{\alpha}$

◇ Deterministic $\rightsquigarrow$ $\bar{\alpha} : (0,\infty) \times \mathcal{Y} \to (0,\infty)$
Statistic $\rightsquigarrow$ $\bar{\alpha} : (0,\infty) \times \mathfrak{M}(\mathbb{M}) \to (0,\infty)$

- We aim for 'convergence' w.r.t. a **loss** $d : \mathfrak{B} \times \mathfrak{B} \to [0,\infty)$ with $d(u,u) = 0$ for all $u \in \mathfrak{B}$
- Typical examples: Bregman distance

$$d\left(u,u^\dagger\right) = \mathcal{D}_{\mathcal{R}}^{u^*}\left(u,u^\dagger\right) := \mathcal{R}(u) - \mathcal{R}\left(u^\dagger\right) - \left\langle u^*, u - u^\dagger\right\rangle$$

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

Regularization schemes

Deterministic Regularization Scheme

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

$$\lim_{\delta \searrow 0} \sup \left\{ d\left(R_{\bar{\alpha}\left(\delta, g^{\mathrm{obs}}\right)}\left(g^{\mathrm{obs}}\right), u^\dagger\right) \;\middle|\; g^{\mathrm{obs}} \in \mathcal{Y}, \|g^{\mathrm{obs}} - F(u^\dagger)\| \leq \delta \right\} = 0$$

Statistical Regularization Scheme

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

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

where G_t is a scaled Poisson process with intensity $F\left(u^\dagger\right)$.

Convergence rates

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

Deterministic convergence rates

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

$$d\left(R_{\bar{\alpha}\left(\delta, g^{\mathrm{obs}}\right)}\left(g^{\mathrm{obs}}\right), u^\dagger\right) = \mathcal{O}\left(\psi\left(\delta\right)\right), \qquad \delta \searrow 0$$

for all $u^\dagger \in M$ and $\left\|g^{\mathrm{obs}} - F\left(u^\dagger\right)\right\|_{\mathcal{Y}} \leq \delta$.

Statistical convergence rates

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

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

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

Outline

Regularization by projection

- Suppose $F = T$ is bounded, linear and positive definite, for simplicity $\mathcal{X} = \mathcal{Y} = \mathbf{L}^2(\mathbb{M})$.
- Regularization by projection: $V_n \subset \mathcal{X}$ with $\dim(V_n) < \infty$ and

$$u_n^{\mathrm{proj}} := \operatorname*{argmin}_{u \in V_n} \left\| Tu - g^{\mathrm{obs}} \right\|_{\mathcal{Y}}^2 \qquad (1)$$

- If $\{v_1, \ldots, v_n\}$ is an orthonormal basis of V_n, then

$$u_n^{\mathrm{proj}} \in V_n : \qquad \left\langle Tu_n^{\mathrm{proj}}, v_j \right\rangle = \int_{\mathbb{M}} v_j g^{\mathrm{obs}} \,\mathrm{d}x, \qquad 1 \leq j \leq n.$$

- Define u_n^{proj} also in the statistical case by replacing g^{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^{proj} and stability.

Some (simplified) results

$$u_n^{\mathrm{proj}} \in V_n : \qquad \left\langle T u_n^{\mathrm{proj}}, v_j \right\rangle = \int_{\mathbb{M}} v_j \,\mathrm{d} G_t, \qquad 1 \leq j \leq n.$$

- 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^{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^{proj} yields a statistical regularization scheme.

Variational regularization I

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

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

  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:

$$u_\alpha \in \operatorname*{argmin}_{u \in \mathfrak{B}} \left[\mathcal{S}\left(G_t; F\left(u\right)\right) + \alpha \mathcal{R}\left(u\right)\right]$$

  where $\mathcal{R}$ is a convex penalty term and $\alpha > 0$ a regularization parameter.

Variational regularization II

$$u_\alpha \in \underset{u \in \mathfrak{B}}{\operatorname{argmin}} \left[\mathcal{S}\left(G_t; F\left(u\right)\right) + \alpha \mathcal{R}\left(u\right)\right]$$

- Main issue in the analysis: data fidelity term lacks of a triangle-type inequality!
- References for deterministic regularization properties: Eggermont & LaRiccia 1996, Resmerita & Anderssen 2007, Pöschl 2007, Bardsley & Laobeul 2008, Bardsley & Luttman 2009, Bardsley 2010, Flemming 2010 & 2011, Lorenz & Worliczek 2013 ...
- 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:

$$\mathcal{S}_0\left(G_t; g\right) = \int_{\mathbb{M}} g \,\mathrm{d}x - \int_{\mathbb{M}} \ln\left(g\right) \mathrm{d}G_t, \qquad g \geq 0 \text{ a.e.}$$

- ideal data misfit functional for exact data $g^\dagger$ given by

$$\mathbf{E}\left[\mathcal{S}_0\left(G_t; g\right)\right] - \mathbf{E}\left[\mathcal{S}_0\left(G_t; g^\dagger\right)\right] = \int_{\mathbb{M}} \left[g - g^\dagger - g^\dagger \ln\left(\frac{g}{g^\dagger}\right)\right] \mathrm{d}x$$

  which is the Kullback-Leibler divergence $\mathbb{KL}\left(g^\dagger; g\right)$.

- we introduce a shift $\sigma > 0$ and consider

$$\mathcal{S}_\sigma\left(G_t; g\right) := \int_{\mathbb{M}} g \,\mathrm{d}x - \int_{\mathbb{M}} \ln\left(g + \sigma\right) \left(\mathrm{d}G_t + \sigma \mathrm{d}x\right)$$

$$\mathcal{T}\left(g^\dagger; g\right) := \mathbb{KL}\left(g^\dagger + \sigma; g + \sigma\right)$$

Assumptions

Assumptions on the problem

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

Assumptions on the method

- $\mathcal{R} : \mathfrak{B} \to (-\infty, \infty]$ is convex, proper and $\tau_{\mathcal{X}}$-sequentially lower semicontinuous.
- $\mathcal{R}$-sublevelsets $\left\{u \in \mathcal{X} \;\middle|\; \mathcal{R}(u) \leq M\right\}$ are $\tau_{\mathcal{X}}$-sequentially pre-compact.

Statistical regularization properties

$$u_\alpha \in \operatorname*{argmin}_{u \in \mathfrak{B}} \left[\mathcal{S}_\sigma\left(G_t; F(u)\right) + \alpha \mathcal{R}(u)\right]$$

Under those assumptions, a minimizer u_α exists with probability one.

Regularization properties (Hohage, W. 2014)

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

$$\lim_{t\to\infty} \bar{\alpha}(t, G_t) = 0, \qquad \lim_{t\to\infty} \frac{\ln(t)}{\sqrt{t}\bar{\alpha}(t, G_t)} = 0$$

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:

$$\beta \mathcal{D}_{\mathcal{R}}^{u^*}\left(u, u^\dagger\right) \leq \mathcal{R}(u) - \mathcal{R}\left(u^\dagger\right) + \varphi\left(\mathcal{T}\left(g^\dagger; F(u)\right)\right) \qquad (2)$$

  for all $u \in \mathfrak{B}$ with $\beta > 0$. φ is assumed to fulfill
  - $\varphi(0) = 0$,
  - $\varphi \nearrow$,
  - φ concave.
- Now the source set $M = M_{\mathcal{R}}^{\varphi}(\beta)$ consists of all $u^\dagger \in \mathfrak{B}$ satisfying (2).

Statistical convergence rates

A priori convergence rates (W., Hohage 2012)

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

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

Under suitable assumptions α can be chosen according to a Lepskĭ-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

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!