Convergence rates in expectation for Tikhonov-type regularization of Inverse Problems with Poisson data

Frank Werner

Institute for numerical and applied mathematics
University of Göttingen, Germany

joint work with Thorsten Hohage

DMV 2012

Outline

Outline

1. Introduction
2. Results on Poisson processes
3. Deterministic convergence analysis
4. Convergence rates in expectation
5. Conclusion

Problem setup

Measurements: Total number N and positions $x_i \in \mathbb{M}$ of photons distributed due to a unknown photon density $g^\dagger$.

Task: Determine the reason $u^\dagger$ of the photon density $g^\dagger$.

Note: The total number N of counted photons depends on the intensity of $g^\dagger$ as well as a parameter t interpreted as exposure time.

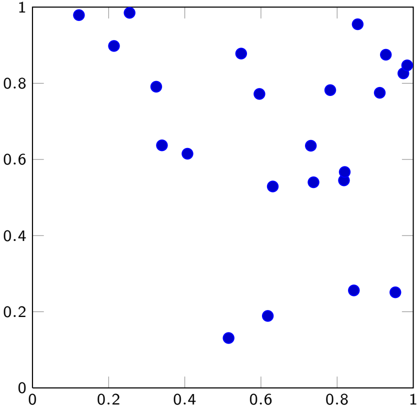
Poisson Processes

Mathematical model:

Poisson Process, i.e.

$$\tilde{G}_t = \sum_{i=1}^{N} \delta_{x_i}$$

with the following properties:



Poisson process - Axiom I

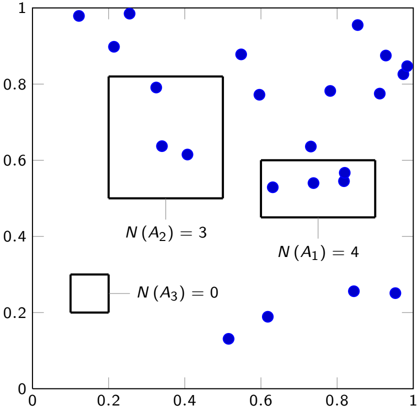
$$N(A) := \#\left\{ i \in \{1, \dots, N\} \;\middle|\; x_i \in A \right\}$$

Independence:

For any choice of $A_1, \dots, A_n \subset \mathbb{M}$ disjoint and measurable, the random variables

$$N(A_1), \dots, N(A_n)$$

are independent.



Poisson process - Axiom II

$$N(A) := \# \left\{ i \in \{1, ..., N\} \;\middle|\; x_i \in A \right\}$$

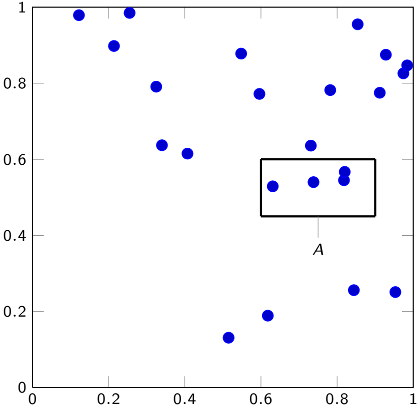
Poisson distribution:

For any measurable $A \subset \mathbb{M}$ the random variable

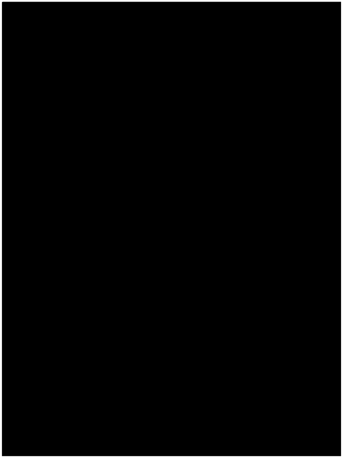
$$N(A)$$

is Poisson distributed with parameter

$$t \int_A g^\dagger \, \mathrm{d}x.$$



Influence of t



We expect 20.000 photons per second

Difficulties I

Model assumption: The imaging process can be descibed by an operator equation

$$F\left(u^{\dagger}\right) = g^{\dagger}$$

where $F : \mathfrak{B} \subset \mathcal{X} \to \mathcal{Y}$ is in general nonlinear and $\mathcal{X}$ and $\mathcal{Y}$ are Banach spaces.

The exact right-hand side $g^{\dagger}$ is unknown and in general F^{-1} is not continuous.

$\Rightarrow$ direct reconstruction impossible, regularization necessary!

Difficulties II

Several applications yield only data for small t, i.e.

- positron emission tomography (radiation exposure)
- astronomical imaging (limited observation time, motion artifacts)
- fluorescence microscopy (photobleaching)

$\Rightarrow$ use a negative log-likelihood aproach to use the infomation at hand on the Poisson distribution:

Minimize

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

over all admissible u.

Approach

We consider a possibly nonlinear,

ill-posed problem with Poisson data.

A. Antoniadis and J. Bigot.
Poisson inverse problems.
Ann. Statist., 34(5):2132–2158, **2006**.

J. M. Bardsley.
A Theoretical Framework for the Regularization of Poisson Likelihood Estimation Problems.
Inverse Probl. Imag., 4:11–17, **2010**.

Tackle the problem with Tikhonov-type regularization:

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

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

Outline

Data fidelity terms

Scaled data $G_t = \frac{1}{t} \sum_{i=1}^{N} \delta_{x_i}$, $tG_t = \tilde{G}_t$ Poisson process.

- Negative log-likelihood:

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

- It holds $\mathbf{E}\left[\mathcal{S}\left(G_t; g\right)\right] = \int_{\mathbb{M}} \left[g - g^\dagger \ln\left(g\right)\right] \, \mathrm{d}x$
- $\rightsquigarrow$ ideal data misfit functional for exact data $g^\dagger$ given by

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

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

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

Controlling the error I

We obtain the following concentration inequality based on

P. Reynaud-Bouret.
Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities.
Probab. Theory Rel., 126(1):103–153, 2003.

Uniform concentration inequality (W., Hohage 2012)

- $\mathbb{M} \subset \mathbb{R}^d$ bounded and Lipschitz,
- $B_s(R) := \left\{ \mathfrak{g} \in H^s(\mathbb{M}) \;\middle|\; \|\mathfrak{g}\|_{H^s(\mathbb{M})} \leq R \right\}$ with $s > d/2, R > 1$.

Then there exists $C_{\mathrm{conc}} = C_{\mathrm{conc}}\left(\mathbb{M}, s, g^\dagger\right) \geq 1$ such that

$$\mathbf{P}\left[\sup_{\mathfrak{g} \in B_s(R)} \left| \int_{\mathbb{M}} \mathfrak{g}\left(\mathrm{d}G_t - g^\dagger \,\mathrm{d}x\right)\right| \leq \frac{\rho}{\sqrt{t}}\right] \geq 1 - \exp\left(-\frac{\rho}{RC_{\mathrm{conc}}}\right)$$

for all $t \geq 1$ and $\rho \geq RC_{\mathrm{conc}}$.

Controlling the error II

- Concentration inequality requires $\mathfrak{g} \in H^s(\mathbb{M}) \subset \mathbf{L}^\infty(\mathbb{M})$ due to $s > d/2$
- Error at $g = F(u)$ leads to $\mathfrak{g} = \ln(F(u))$

$\Rightarrow$ Too strong assumption!

$\rightsquigarrow$ Shift by $\sigma > 0$:

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

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

- Then the error is given by

$$\left| \int_{\mathbb{M}} \ln(g + \sigma) \left(\mathrm{d}G_t - g^\dagger \,\mathrm{d}x\right) \right| .$$

Outline

Noise level I

We have two data fidelity terms:

- $\mathcal{S}$ w.r.t. the measured data g^{obs}
- $\mathcal{T}$ w.r.t. the photon density $g^{\dagger}$

As before: consider the difference between both as noise level!

Noise level

There exist constants **err** ≥ 0 and $C_{\mathrm{err}} \geq 1$ such that

$$\mathcal{S}\left(g^{\mathrm{obs}}; g\right) - \mathcal{S}\left(g^{\mathrm{obs}}; g^{\dagger}\right) \geq \frac{1}{C_{\mathrm{err}}}\mathcal{T}\left(g^{\dagger}; g\right) - \mathbf{err}$$

for all $g \in F(\mathfrak{B})$.

Noise level II

- *Classical deterministic noise model:*
  If $\mathcal{S}(g;\hat{g}) = \mathcal{T}(g;\hat{g}) = \|g-\hat{g}\|_{\mathcal{Y}}^{r}$, then $C_{\mathrm{err}} = 2^{r-1}$ and $\mathbf{err} = 2\left\|g^{\dagger} - g^{\mathrm{obs}}\right\|_{\mathcal{Y}}^{r}$.
- *Poisson data:*
  $C_{\mathrm{err}} = 1$ and

$$\mathbf{err} \geq -\int\limits_{\mathbb{M}} \ln\left(g^{\dagger}+\sigma\right)\left(\mathrm{d}G_t - g^{\dagger}\,\mathrm{d}x\right) + \int\limits_{\mathbb{M}} \ln\left(F(u)+\sigma\right)\left(\mathrm{d}G_t - g^{\dagger}\,\mathrm{d}x\right)$$

for all $u \in \mathfrak{B}$.
Uniform concentration inequality: $\mathbf{err} \leq \frac{2\rho}{\sqrt{t}}$ with probability $\geq 1-\exp(-c\rho)$ for some constant $c > 0$.

Source condition I

- Bregman distance:

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

where $u^* \in \partial\mathcal{R}\left(u^\dagger\right) \subset \mathcal{X}'$.

- Use a variational inequality as source condition:

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

for all $u \in \mathfrak{B}$ with $\beta > 0$. φ is assumed to fulfill
  - $\varphi\left(0\right) = 0$,
  - $\varphi \nearrow$,
  - φ concave.

Source condition II

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

- does not depend on the structure of $\mathcal{X}$ and $\mathcal{Y}$
- nonlinear F: combination of source and nonlinearity condition

$\mathcal{X}$, $\mathcal{Y}$ Hilbert spaces, $\mathcal{R}(u) = \|u - u_0\|_{\mathcal{X}}^2$:

- $F(\mathfrak{B}) \subset \mathbf{L}^\infty$ bounded: spectral source + nonlinearity condition imply variational inequality (use $\left\|F(u) - g^\dagger\right\|_{\mathbf{L}^2}^2 \leq C\mathcal{T}\left(g^\dagger; F(u)\right)$)

J. M. Borwein and A. S. Lewis.
Convergence of best entropy estimates.
SIAM J. Optimization, 1:191–205, 1991.

- $\mathcal{T}(g_2; g_1) = \|g_1 - g_2\|_{\mathcal{Y}}^2$: obtained convergence rates are optimal

Deterministic convergence analysis I

Suppose

- the noise assumption is fulfilled with $\mathbf{err} \geq 0$ and
- the variational inequality holds true.

Theorem (error decomposition)

Then

$$\beta \mathcal{D}_{\mathcal{R}}^{u^*}\left(u_\alpha, u^\dagger\right) \leq \frac{\mathbf{err}}{\alpha} + (-\varphi)^*\left(-\frac{1}{C_{\mathrm{err}}\alpha}\right)$$

for all $\alpha > 0$.

Fenchel conjugate:

$$(-\varphi)^*(s) = \sup_{\tau \geq 0}\left(s\tau + \varphi(\tau)\right).$$

Deterministic convergence analysis II

$$\beta \mathcal{D}_{\mathcal{R}}^{u^*}\left(u_\alpha, u^\dagger\right) \leq \frac{\mathbf{err}}{\alpha} + (-\varphi)^*\left(-\frac{1}{C_{\mathrm{err}}\alpha}\right)$$

Theorem (a priori rates)

The infimum of the right-hand side it attained at $\alpha = \overline{\alpha}$ if and only if

$$\frac{-1}{C_{\mathrm{err}}\overline{\alpha}} \in \partial(-\varphi)(C_{\mathrm{err}}\,\mathbf{err}) \qquad \left[\hat{=} \quad \bar{\alpha} = \frac{1}{C_{\mathrm{err}}\varphi'\left(C_{\mathrm{err}}\,\mathbf{err}\right)}\right]$$

and in that case

$$\beta \mathcal{D}_{\mathcal{R}}^{u^*}\left(u_{\overline{\alpha}}, u^\dagger\right) \leq C_{\mathrm{err}}\varphi\left(\mathbf{err}\right).$$

Deterministic convergence analysis III

Suppose moreover $\mathcal{X}$ Hilbert space, $\mathcal{R}(u) = \|u - u_0\|_{\mathcal{X}}^2$, $\beta \geq \frac{1}{2}$. Set

- $r > 1$
- $\alpha_j := \mathbf{err}\, r^{2j-2}$ for $j = 2, ..., m$ such that $\alpha_{m-1} < 1 \leq \alpha_m$
- $j_{\mathrm{bal}} := \max \left\{ j \leq m \;\middle|\; \left\| u_{\alpha_i} - u_{\alpha_j} \right\|_{\mathcal{X}} \leq 4\sqrt{2} r^{1-i} \text{ for all } i < j \right\}$

Theorem (a posteriori rates)

Then for $\mathbf{err} > 0$ sufficiently small:

$$\left\| u_{\alpha_{j_{\mathrm{bal}}}} - u^\dagger \right\|_{\mathcal{X}}^2 \leq 6r \min_{j=1,\ldots,m} \left[\frac{\mathbf{err}}{\alpha_j} + (-\varphi)^* \left(-\frac{1}{C_{\mathrm{err}} \alpha_j} \right) \right].$$

If $\varphi^{1+\varepsilon}$ is additionally concave ($\varepsilon > 0$), then

$$\left\| u_{\alpha_{j_{\mathrm{bal}}}} - u^\dagger \right\|_{\mathcal{X}}^2 \leq 6r^{1+\frac{1}{\varepsilon}} C_{\mathrm{err}} \varphi(\mathbf{err})$$

as $\mathbf{err} \searrow 0$.

Outline

Convergence rates for known φ

Suppose

- $\mathcal{X}$ Banach space, $u^\dagger \in \mathfrak{B} \subset \mathcal{X}$ bounded, closed and convex
- $\mathbb{M} \subset \mathbb{R}^d$ bounded and Lipschitz
- $F(u) \geq 0$ a.e. for all $u \in \mathfrak{B}$
- there exists a Sobolev index $s > \frac{d}{2}$ such that $F(\mathfrak{B})$ is a bounded subset of $H^s(\mathbb{M})$

A priori convergence rates (W., Hohage 2012)

Then for $\alpha = \alpha(t)$ such that $\frac{1}{\alpha} \in -\partial(-\varphi)\left(\frac{1}{\sqrt{t}}\right)$ we obtain the convergence rate

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

Convergence rates for unknown φ

Suppose moreover $\mathcal{X}$ Hilbert space, $\mathcal{R}(u) = \|u - u_0\|_{\mathcal{X}}^2$, $\beta \geq \frac{1}{2}$, $\varphi^{1+\varepsilon}$ concave ($\varepsilon > 0$). Set

- $r > 1$, $\tau > 0$ sufficiently large
- $\alpha_j := \frac{\tau \ln(t)}{\sqrt{t}} r^{2j-2}$ for $j = 2, ..., m$ such that $\alpha_{m-1} < 1 \leq \alpha_m$
- $j_{\mathrm{bal}} := \max \left\{ j \leq m \;\middle|\; \left\|u_{\alpha_i} - u_{\alpha_j}\right\|_{\mathcal{X}} \leq 4\sqrt{2} r^{1-i} \text{ for all } i < j \right\}$

A posteriori convergence rates (W., Hohage 2012)

Then we obtain

$$\mathbf{E}\left[\left\|u_{\alpha_{j_{\mathrm{bal}}}} - u^\dagger\right\|_{\mathcal{X}}^2\right] = \mathcal{O}\left(\varphi\left(\frac{\ln(t)}{\sqrt{t}}\right)\right) \qquad \text{as} \qquad t \to \infty.$$

Adaptivity causes a loss of a logarithmic factor!

A. Tsybakov.
On the best rate of adaptive estimation in some inverse problems.
C. R. Acad. Sci. Paris, 330:835–840, 2000.

Outline

Presented results

- Improvements in the theory of inverse problems with Poisson data:
  - convergence and convergence rates
  - generalized source conditions
  - a priori and a posteriori parameter choice
- regularization theory with general data fidelity terms

F. Werner and T. Hohage.
Convergence rates in expectation for Tikhonov-type regularization of Inverse Problems with Poisson data.
Inverse Problems, to appear, 2012.

Thank you for your attention!