Regularization of statistical inverse problems and applications

Thorsten Hohage

Institut für Numerische und Angewandte Mathematik
Georg-August Universität Göttingen

September 22-25, 2015

Outline

inverse problems setup

- $\mathcal{X}, \mathcal{Y}$ Banach spaces, $\mathrm{dom}(F) \subset \mathcal{X}$ closed, convex
- $F : \mathrm{dom}(F) \to \mathcal{Y}$ continuous, injective forward operator
- $u^\dagger \in \mathrm{dom}(F)$ unknown quantity, $g^\dagger := F(u^\dagger)$ observable quantity, g^{obs} observations/measurements of $g^\dagger$
- ill-posedness: Typically F^{-1} not continuous

aims:

- Construct "good" estimators $\widehat{u}_\alpha$ of $u^\dagger$ given g^{obs}
- Prove convergence of $\widehat{u}_\alpha$ to $u^\dagger$ as g^{obs} tends to $g^\dagger$ in some sense

Examples of noise models

- finite dimensional noise models
- white noise models $\rightsquigarrow$ Lecture 1
- point processes, e.g. Poisson processes
- impulsive noise (only deterministic so far)

In all cases (except possibly the last one) we have $g^{\mathrm{obs}} \notin \mathcal{Y}$!

data fidelity functional

- $\|g^{\mathrm{obs}} - F(u)\|_{\mathcal{Y}}$ not well defined if $g^{\mathrm{obs}} \notin \mathcal{Y}$!
- We assume that we can associate with each g^{obs} a convex data fidelity functional

$$\mathcal{S}_{g^{\mathrm{obs}}}(\cdot) : \mathcal{Y} \to (-\infty, \infty]$$

  such that $\mathcal{S}_{g^{\mathrm{obs}}}(g^{\dagger}) \approx \inf_{g \in \mathcal{Y}} \mathcal{S}_{g^{\mathrm{obs}}}(g)$.
- We only access the data g^{obs} via its associated data fidelity functional $\mathcal{S}_{g^{\mathrm{obs}}}(\cdot)$.
- natural choice of $\mathcal{S}$: negative log-likelihood functional

$$\mathcal{S}_{g^{\mathrm{obs}}}(g) = -\ln \mathbb{P}_g[g^{\mathrm{obs}}] + C$$

  with C independent of g.

variational regularization

$$\widehat{u}_\alpha \in \operatorname{argmin}_{u \in \mathrm{dom}(F)} \left[\mathcal{S}_{g^{\mathrm{obs}}}(F(u)) + \alpha\mathcal{R}(u)\right]$$

examples of penalty functionals:

- $\mathcal{R}(u) = \|u - u_0\|_{\mathcal{X}}^p$ for some $p \geq 1$
- $\mathcal{R}(u) = \sum_j |\langle u, v_j \rangle|$ for some ONB or frame $\{v_j\}$
- $\mathcal{R}(u) = |u|_{TV}$
- $\mathcal{R}(u) = \int u(x) \ln u(x)\,\mathrm{d}x$ (entropy regularization)

effective noise level

data fidelity functional for exact data $g^\dagger := F(u^\dagger)$:

- $\mathcal{T}_{g^\dagger} : \mathrm{ran}(F) \to [0,\infty]$
- assumed property: $\mathcal{T}_{g^\dagger}(g) = 0 \;\Leftrightarrow\; g = g^\dagger$
- frequent choice: $\mathcal{T}_{g^\dagger}(g) = \mathbb{E}\left[\mathcal{S}_{g^{\mathrm{obs}}}(g) - \mathcal{S}_{g^{\mathrm{obs}}}(g^\dagger)\right]$

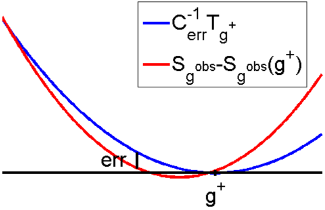
Definition Let $C_{\mathrm{err}} \geq 1$ and $\tilde{\mathcal{Y}} \subset \mathrm{ran}(F)$.

$$\mathrm{err}_{\tilde{\mathcal{Y}}} := \sup_{g\in\tilde{\mathcal{Y}}}\left[-\mathcal{S}_{g^{\mathrm{obs}}}(g) + \mathcal{S}_{g^{\mathrm{obs}}}\left(g^\dagger\right) + \frac{1}{C_{\mathrm{err}}}\mathcal{T}_{g^\dagger}(g)\right]$$

We call $\mathrm{err} := \mathrm{err}_{\mathrm{ran}(F)}$ (or $\mathrm{err}_{\tilde{\mathcal{Y}}}$) the effective noise level (on $\tilde{\mathcal{Y}}$).

$\mathrm{err}_{\tilde{\mathcal{Y}}}$ is defined s.t. $\forall g \in \tilde{\mathcal{Y}}$

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



estimating the effective noise level

Gaussian white noise

Assume that $g^{\mathrm{obs}} = g^{\dagger} + \xi$ where ξ is Gaussian white noise

- Choose $\mathcal{S}_{g^{\mathrm{obs}}}(g) := \|g\|^2 - 2\left\langle g^{\mathrm{obs}}, g\right\rangle$.
- Choose $\mathcal{T}_{g^{\dagger}}(g) = \mathbb{E}\left[\mathcal{S}_{g^{\mathrm{obs}}}(g) - \mathcal{S}_{g^{\mathrm{obs}}}\left(g^{\dagger}\right)\right] = \|g - g^{\dagger}\|^2$
- For $C_{\mathrm{err}} = 1$ we get

$$\mathrm{err}_{\tilde{\mathcal{Y}}} = 2 \sup_{g \in \tilde{\mathcal{Y}}} \left\langle \xi, g - g^{\dagger} \right\rangle .$$

- Concentration inequalities for err are a well-studied in probability theory.

estimating the effective noise level

standard deterministic noise model

If $\|g^{\mathrm{obs}} - g^{\dagger}\|_{\mathcal{Y}} \leq \delta$ and

$$\mathcal{S}_{g_1}(g_2) = \mathcal{T}_{g_1}(g_2) = \|g_1 - g_2\|_{\mathcal{Y}}^{p},$$

then the effective noise level on $\mathcal{Y}$ with $C_{\mathrm{err}} = 2^{p-1}$ is bounded by

$$\mathrm{err} \leq 2\delta^{p}.$$

estimating the effective noise level

discrete noise model

noise model: $g_i^{\mathrm{obs}} = g^\dagger(x_i) + \epsilon_i \qquad , i = 1, \dots, n$
quadrature rule: $Q_n\varphi := \sum_{i=1}^n \alpha_i \varphi(x_i) \approx \int_\Omega \varphi(x)\,\mathrm{d}x$
data fidelity functionals:

$$\mathcal{S}_{g^{\mathrm{obs}}}(g) := \|g\|^2 - 2\sum_{i=1}^{n} \alpha_i g_i^{\mathrm{obs}} g(x_i)$$

$$\mathcal{T}_{g^\dagger}(g) := \|g - g^\dagger\|_{L^2(\Omega)}^2$$

effective noise level for $C_{\mathrm{err}} = 1$:

$$\mathrm{err}_{\tilde{\mathcal{Y}}} := \sup_{g\in\tilde{\mathcal{Y}}} \Bigg[\underbrace{(Q_n - I)(g^\dagger(g - g^\dagger))}_{\text{discretization error}} + \underbrace{\sum_{i=1}^{n} \alpha_i \epsilon_i (g(x_i) - g^\dagger(x_i))}_{\text{random error}} \Bigg]$$

Outline

Bregman distances

Definition

Let $\mathcal{X}$ be a Banach space, $\mathcal{R} : \mathcal{X} \to \mathbb{R} \cup \{\infty\}$ convex, $u_1 \in \mathcal{X}$, and $u_1^* \in \partial\mathcal{R}(u_1)$. Then

$$D_{\mathcal{R}}^{u_1^*}(u_2, u_1) := \mathcal{R}(u_2) - \mathcal{R}(u_1) - \langle u_1^*, u_2 - u_1 \rangle$$

is called the Bregman distance of $\mathcal{R}$ at u_1 and u_2.

properties:

- $D_{\mathcal{R}}^{u_1^*}(u_2, u_1) \geq 0$, and $D_{\mathcal{R}}^{u_1^*}(u_1, u_1) = 0$.
- If $\mathcal{R}$ is strictly convex, then $D_{\mathcal{R}}^{u_1^*}(u_2, u_1) = 0$ implies $u_2 = u_1$.
- In a Hilbert space with $\mathcal{R}(u_1) = \|u_1\|^2$ we have $\partial\mathcal{R}(u_1) = \{2u_1^*\}$ and $D_{\mathcal{R}}^{u_1^*}(u_1, u_2) = \|u_1 - u_2\|^2$.

P Eggermont. *Maximum entropy regularization for Fredholm integral equations of the first kind*, **SIAM J. Math. Anal.** 24:1557–1576, 1993

M. Burger, S. Osher. *Convergence rates of convex variational regularization.* **Inverse Problems** 20:1411–1422, 2004.

source conditions

spectral source condition:

$$u^\dagger = \varphi\left(F'[u^\dagger]^* F'[u^\dagger]\right) w$$

variational source condition (VSC): Let $\beta \in (0,1]$.

$$\beta D_{\mathcal{R}}(u, u^\dagger) \leq \mathcal{R}(u) - \mathcal{R}(u^\dagger) + \psi\left(\mathcal{T}_{F(u^\dagger)}(F(u))\right) \qquad \text{for all } u \in \mathrm{dom}(F).$$

First used (with $\psi(t) = c\sqrt{t}$) in

B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. *A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators*. **Inverse Problems** 23:987–1010, 2007.

Here $\varphi, \psi : [0,\infty) \to [0,\infty)$ are non-decreasing and vanish at 0. ψ is assumed be concave.

advantages of variational vs. spectral source conditions

- simpler proofs
- not only sufficient, but even necessary for certain rates of convergence (for linear operators in Hilbert spaces)[1]
- VSCs do not involve F' $\rightsquigarrow$ no need of tangential cone condition or related conditions
- VSC work for Banach spaces and general $\mathcal{R}$, $\mathcal{S}$.

1) see:

J. Flemming, B. Hofmann, and P. Mathé. *Sharp converse results for the regularization error using distance functions.* **Inverse Problems**, 27:025006, 2011.

convergence of Tikhonov regularization

Theorem

Assume VSC and the existence of a global minimizer of the Tikhonov functional.

1. *Let* $(-\psi)^*(s) := \sup_{t\geq 0}[ts + \psi(t)]$ *denote the Fenchel conjugate. Then*

$$\beta D_{\mathcal{R}}(\widehat{u}_\alpha, u^\dagger) \leq \frac{\mathrm{err}}{\alpha} + (-\psi)^* \left(-\frac{1}{C_{\mathrm{err}}\alpha}\right).$$

2. *If we choose* $\frac{-1}{C_{\mathrm{err}}\alpha} \in \partial(-\psi)(C_{\mathrm{err}}\mathrm{err})$, *then*

$$\beta D_{\mathcal{R}}(\widehat{u}_\alpha, u^\dagger) \leq C_{\mathrm{err}}\psi(\mathrm{err}).$$

M. Grasmair. *Generalized Bregman distances and convergence rates for non-convex regularization methods.* **Inverse Problems** 26:115014 (16p.), 2010.

R. I. Bot and B. Hofmann. *An extension of the variational inequality approach for nonlinear ill-posed problems.* **J. Int. Eq. Appl.** 22:369–392, 2010.

J. Flemming. *Theory and examples of variational regularisation with non-metric fitting functionals.* **J. Inv. Ill-Posed Probl.** 18:677–699, 2010.

Proof of convergence theorem, part 1

Proof. By the definition of $\widehat{u}_\alpha$ we have

$$\mathcal{S}_{g^{\mathrm{obs}}}\left(F(\widehat{u}_\alpha)\right) + \alpha\mathcal{R}(\widehat{u}_\alpha) \le \mathcal{S}_{g^{\mathrm{obs}}}\left(F(u^\dagger)\right) + \alpha\mathcal{R}(u^\dagger)\,. \qquad (1)$$

It follows that

$$\begin{aligned}
\beta D_{\mathcal{R}}(\widehat{u}_\alpha, u^\dagger) &\overset{\text{Ass. SC}}{\le} \mathcal{R}(\widehat{u}_\alpha) - \mathcal{R}(u^\dagger) + \psi\left(\mathcal{T}_{F(u^\dagger)}\left(F(\widehat{u}_\alpha)\right)\right) \\
&\overset{(1)}{\le} \frac{1}{\alpha}\left(\mathcal{S}_{g^{\mathrm{obs}}}\left(F(u^\dagger)\right) - \mathcal{S}_{g^{\mathrm{obs}}}\left(F(\widehat{u}_\alpha)\right)\right) \\
&\qquad + \psi\left(\mathcal{T}_{F(u^\dagger)}\left(F(\widehat{u}_\alpha)\right)\right) \\
&\overset{\text{Ass. N}}{\le} \frac{\mathrm{err}}{\alpha} - \frac{1}{C_{\mathrm{err}}\alpha}\mathcal{T}_{F(u^\dagger)}\left(F(\widehat{u}_\alpha)\right) + \psi\left(\mathcal{T}_{F(u^\dagger)}\left(F(\widehat{u}_\alpha)\right)\right) \\
&\le \frac{\mathrm{err}}{\alpha} + \sup_{t\ge 0}\left[\frac{t}{-C_{\mathrm{err}}\alpha} - (-\psi)(t)\right] \\
&= \frac{\mathrm{err}}{\alpha} + (-\psi)^*\left(-\frac{1}{C_{\mathrm{err}}\alpha}\right)\,.
\end{aligned}$$

Proof of convergence theorem, part 2

$$
\begin{aligned}
&\inf_{\alpha>0}\left[\frac{\mathrm{err}}{\alpha}+(-\psi)^*\left(-\frac{1}{C_{\mathrm{err}}\alpha}\right)\right] \\
=&-\sup_{t<0}\left[tC_{\mathrm{err}}\mathrm{err}-(-\psi)^*(t)\right] \\
=&-(-\psi)^{**}(C_{\mathrm{err}}\mathrm{err}) \\
=&\psi(C_{\mathrm{err}}\mathrm{err}) \le C_{\mathrm{err}}\psi(\mathrm{err})\,.
\end{aligned}
$$

By the conditions for equality in Young's inequality, the supremum is attained at α if and only if $\frac{-1}{C_{\mathrm{err}}\alpha} \in \partial(-\psi)(C_{\mathrm{err}}\mathrm{err})$.

selection of the regularization parameter

Note: discrepancy principle not applicable in general.

Lepskiĭ balancing principle: Let $\alpha_j := r^j \mathrm{err}$ with $r > 1$ and choose

$$\alpha_{\mathrm{bal}} := \max\left\{ j \in \mathbb{N} : \|u_{\alpha_j} - u_{\alpha_k}\| \leq 4(4C_{\mathcal{X}} r^{-j})^{\frac{1}{q}} \text{ for } k = 0, \ldots, j-1 \right\}.$$

Theorem

If $\psi^{1+\epsilon}$ is concave for some $\epsilon > 0$ then

$$\|u_{\alpha_{\mathrm{bal}}} - u^\dagger\|^2 \leq C\psi(\mathrm{err})\,.$$

F. Werner, T. Hohage. *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*, **Inverse Problems** 28:104004 (16p.), 2012.

Newton-type methods

Disadvantages of Tikhonov-type regularization: minimization of non-convex functional, no uniqueness of minimizers.
Alternative: Choose $\alpha_k = \alpha_0 \rho^k$ for some $\rho \in (0,1)$ and set

$$u_{k+1} \in \operatorname{argmin}_{u \in \mathrm{dom}(F)} \left[\mathcal{S}_{g^{\mathrm{obs}}} \left(F'[u_k](u - u_k) + F(u_k) \right) + \alpha_k \mathcal{R}(u) \right]$$

If $\mathcal{S}$ and $\mathcal{R}$ are convex, a convex optimization problem has to be solved in each Newton step. We use an algorithm from

A. Chambolle, T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging.* J. Math. Imaging Vis 40:120-145, 2011.

T. Hohage, C. Homann. A Generalization of the Chambolle-Pock algorithm to Banach spaces with Applications to Inverse Problems. arXiv 1412.0126, 2014.

Under an additional assumption on the local approximation quality of F' (a tangential cone condition) we can show similar results as for Tikhonov regularization.

T. Hohage, F. Werner. *Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data*. **Numer. Math.**, **Numer. Math.** 123:745–779, 2013.

selected references

- T. Hohage, F. Werner. *Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data*. **Numer. Math.**123:745–779, 2013.
- F. Werner, T. Hohage. *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*, **Inverse Problems** 28:104004 (16p.), 2012.
- T. Schuster, B. Kaltenbacher, B. Hofmann, K. Kazimierski. *Regularization Methods in Banach Spaces* In: Radon Series on Computational and Applied Mathematics, deGruyter, 2012.
- B. Kaltenbacher, B. Hofmann. *Convergence rates for the iteratively regularized Gauss-Newton method in Banach spaces*, **Inverse Problems** 26:035007 (21p.), 2010.
- O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, F. Lenzen. *Variational methods in imaging.* Springer, 2009

Outline

photonic imaging

- Neglecting read-out errors, photon data can be considered as realizations of a Poisson process.
- After binning data consist of vectors or arrays of independent Poisson distributed integers.
- examples:
  - coherent x-ray imaging (with Tim Salditt)
  - scanning fluorescence microscopy, e.g. standard confocal, 4Pi or STED microscopy (with Stefan Hell, Alexander Egner)
  - Positron Emission Tomography (PET)
  - astronomical imaging

point processes

A point process on submanifold $\mathbb{M} \subset \mathbb{R}^d$ can either be defined as

- a random finite set of points $\{x_1, \ldots, x_N\} \subset \mathbb{M}$
- or as a finite sum of Dirac measures: $$Y = \sum_{i=1}^{N} \delta_{x_i} \, .$$

In general N is random.

A point process is called Poisson process with density $g \in L^1(\mathbb{M})$, $g \geq 0$ if:

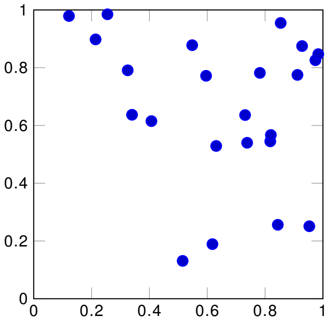


Figure: simulated point process

Poisson processes: first defining property

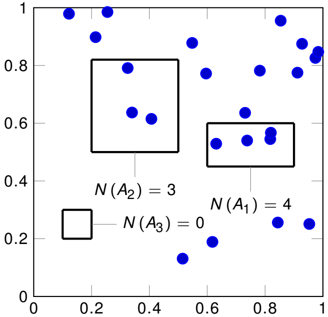
$$N(A) := \#\{i \in \{1, ..., N\} \mid x_i \in A\}$$

1. independence

For any disjoint, measurable subsets $A_1, ..., A_n \subset \mathbb{M}$ the random numbers

$$N(A_1), ..., N(A_n)$$

are independent.



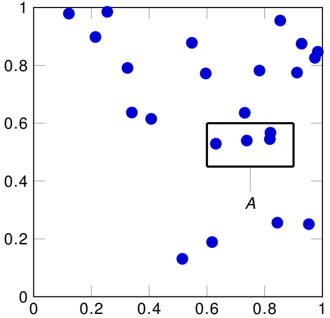
Poisson processes: second defining property

$$N(A) := \#\{i \in \{1, ..., N\} \mid x_i \in A\}$$

2. expectation= intensity
For any measurable $A \subset \mathbb{M}$

$$\mathbb{E}[N(A)] = \int_A g(x)\,\mathrm{d}x\,.$$

Then $N(A)$ can be shown to be Poisson distributed with parameter $\lambda = \int_A g(x)\,\mathrm{d}x$, i.e.
$\mathbb{P}[N(A) = n] = \exp(-\lambda)\frac{\lambda^n}{n!}$.



properties of Poisson processes

Writing the Poisson process as $Y = \sum_{j=1}^{N} \delta_{x_j}$ we have

$$\int \psi dY = \sum_{j=1}^{N} \psi(x_j)$$

If Y has density $g^\dagger$ and $\psi : \mathbb{M} \to \mathbb{R}$ is measurable, then

$$\mathbb{E}\left[\int_{\mathbb{M}} \psi dY\right] = \int_{\mathbb{M}} \psi g^\dagger \, dx$$
$$\mathbf{Var}\left[\int_{\mathbb{M}} \psi dY\right] = \int_{\mathbb{M}} \psi^2 g^\dagger \, dx$$

whenever the integrals on the rhs exist.

log-likelihood

- negative log-likelihood (with scaling factor $1/t$, up to additive constants)

$$\mathcal{S}_{Y_t}(g) = \begin{cases} \int_{\mathbb{M}} g\,dx - \int_{\mathbb{M}} \ln(g)\,\mathrm{d}Y_t = \int_{\mathbb{M}} g\,dx - \sum_{j=1}^{N} \frac{1}{t}\ln(g(x_j)), \\ \infty, \end{cases}$$

- expectation: (convention $\ln(x) := -\infty$ for $x \leq 0$)

$$\mathbb{E}\left[\mathcal{S}_{Y_t}(g)\right] = \int_{\mathbb{M}} \left[g - g^{\dagger}\ln(g)\right]\,dx\,.$$

- Kullback-Leibler divergence
$\mathcal{T}_{g^{\dagger}}(g) = \mathbb{E}\left[\mathcal{S}_{Y_t}(g) - \mathcal{S}_{Y_t}\left(g^{\dagger}\right)\right] = \mathrm{KL}\left(g^{\dagger}; g\right)$ where

$$\mathrm{KL}\left(g^{\dagger}; g\right) := \int_{\mathbb{M}} \left[g - g^{\dagger} - g^{\dagger}\ln\left(\frac{g}{g^{\dagger}}\right)\right]\,dx\,.$$

setup for Poisson data

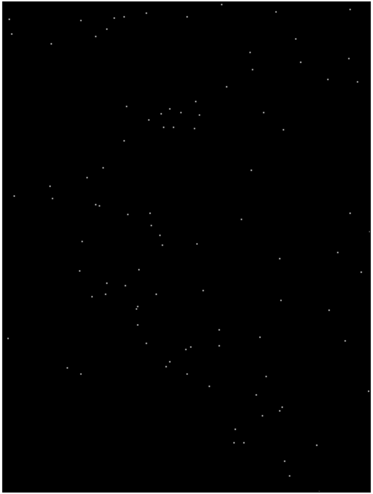
- $\mathcal{X}$ Banach space, $\mathcal{Y} = L^1(\mathbb{M}) \cap L^\infty(\mathbb{M})$.
- $F : \mathrm{dom}(F) \to \mathcal{Y}$ satisfies $F(u) \geq 0$ for all $u \in \mathrm{dom}(F)$.
- tY_t, $t > 0$ Poisson process with intensity $tg^\dagger$.
- $\mathcal{T}_{g^\dagger}(g) := \mathrm{KL}\left(g^\dagger + \sigma; g + \sigma\right)$ with small $\sigma > 0$.
- $\mathcal{S}_{g^\dagger}(g) := \int_{\mathbb{M}} g\,dx - \int_{\mathbb{M}} \ln(g + \sigma)(dY_t + \sigma dx)$

t can often be interpreted as exposure time and is proportional to the expected number of photons.

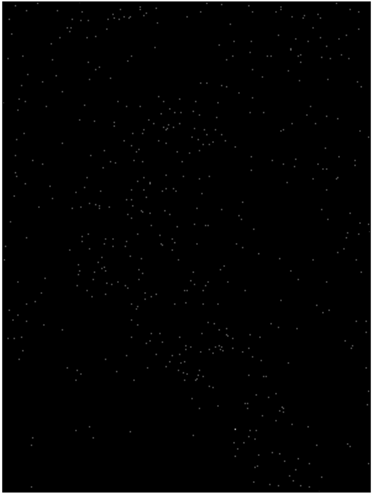
aims:

- Prove convergence as $t \to \infty$
  or answer the question
- How much can we learn from N photons?

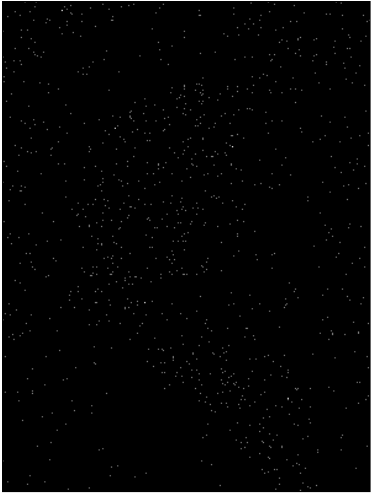
Y_t with $t = 100$ expected points



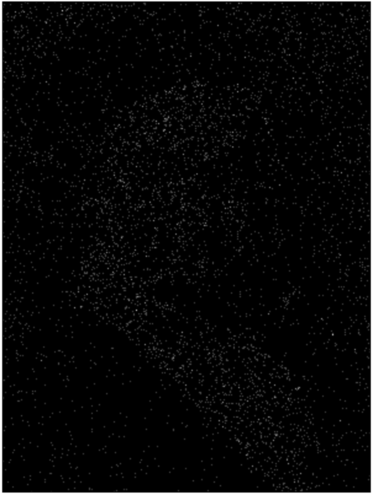
Y_t with $t = 500$ expected points



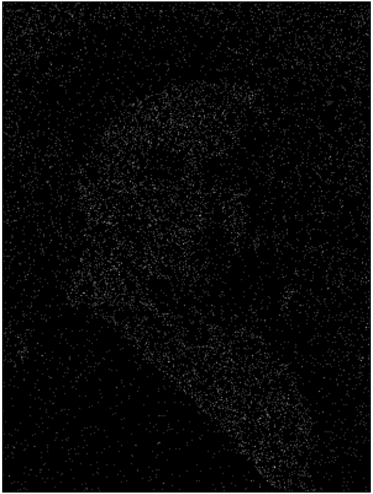
Y_t with $t = 1000$ expected points



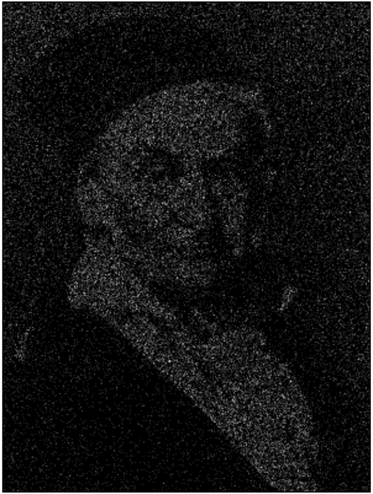
Y_t with $t = 5000$ expected points



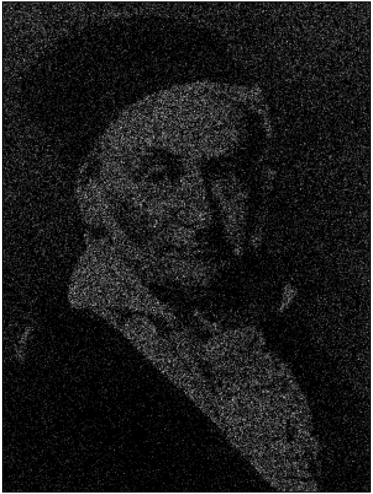
Y_t with $t = 10.000$ expected points



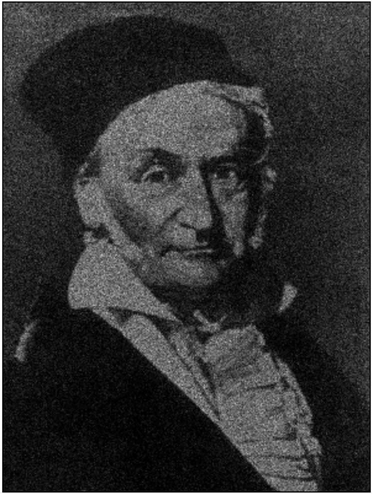
Y_t with $t = 50.000$ expected points



Y_t with $t = 100.000$ expected points



Y_t with $t = 1.000.000$ expected points



the true intensity $g^\dagger$

effective noise level

$$\mathrm{err}(g) = \int_{\mathbb{M}} \frac{g+\sigma}{g^{\dagger}+\sigma}\,(dY_t - g^{\dagger}dx)$$

- uniform boundedness of integrands essential for concentration inequalities
- need offset $\sigma > 0$
- σ can be chosen very small ($= 10^{-4}$ or 10^{-6}), only logarithmic dependence on σ

a concentration inequality

Proposition Let $\mathcal{Y} = L^1(\mathbb{M})$ with $\mathbb{M} \subset \mathbb{R}^d$ a bounded Lipschitz domain. Assume that $F(u) \geq 0$ for all $u \in \mathrm{dom}(F)$ and

$$\sup_{u \in \mathrm{dom}(F)} \|F(u)\|_{H^s} < \infty \qquad \text{for some } s > d/2.$$

Then there exists $C > 0$ such that

$$\mathbb{P}\left[\mathrm{err} \geq \frac{\rho}{\sqrt{t}}\right] \leq \exp\left(-\frac{\rho}{C}\right), \qquad \forall t, \rho \geq 1.$$

Proof based on

P. Reynaud-Bouret. *Adapative estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities.* **Prob. Theory Rel.** 126:103–153, 2003.

P. Massart. *About the constants in Talagrand's concentration inequalities for empirical processes.* **Ann.Prob.** 28:863–884, 2000.

M. Talagrand. *New concentration inequalities in product spaces.* **Invent. Math.** 126:505–563, 1996.

convergence in expectation for Poisson data

Corollary

Under the assumptions of the previous proposition and Assumption SC generalized Tikhonov regularization with a-priori parameter choice rule $\frac{-1}{\alpha} \in \partial(-\psi)(t^{-1/2})$ *fulfills the error estimate*

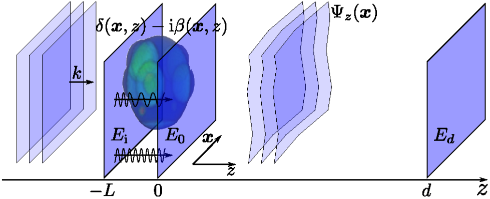
$$\mathbb{E}\left[\|\widehat{u} - u^\dagger\|^2\right] = \mathcal{O}\left(\psi\left(t^{-1/2}\right)\right) \qquad t \to \infty.$$

A similar result holds for a Lepskiĭ stopping rule, but we loose a logarithmic factor in t.

F. Werner, T. Hohage. *Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data*, **Inverse Problems** 28:104004 (16p.), 2012.

Outline

phase retrieval in optics



refractive index: $n(\mathbf{x},z) = 1-\delta(\mathbf{x},z)+i\beta(\mathbf{x},z)$, $0 \le \beta,\delta \ll 1$

unknown: $u(\mathbf{x}) = \kappa\int_{-\infty}^{0}[\delta(\mathbf{x},z) - i\beta(\mathbf{x},z)]\,\mathrm{d}z$

forward operator: $F_\gamma : L^s(B_\rho,\mathbb{C}) \longrightarrow L^1(\mathbb{R}^2)$ parameterized by dimensionless Fresnel number $\gamma > 0$ proportional to $1/d$:

$$(F_\gamma(u))(x) := \left|\int_{B_\rho} \exp(i\gamma|x-y|^2)e^{\mathrm{i}u(y)}\,\mathrm{d}y\right|^2$$

far-field case: Limit $\gamma \to 0$ or $d \to \infty$ (after a rescaling)

$$(F_0(u))(x) := \left|\int_{B_\rho} \exp(ix\cdot y)e^{\mathrm{i}u(y)}\,\mathrm{d}y\right|^2.$$

Theorem

For all $\gamma > 0$ the operator F_γ is injective.

- Surprising since F_γ maps complex to real images.
- Only assumption: Compactness of supp(u)
- Proof relies on theory of entire functions

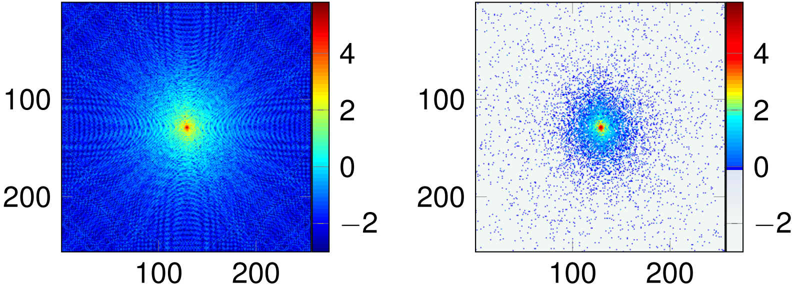
S. Maretzke *A uniqueness result for propagation-based phase contrast imaging from a single measurement.* **Inverse Problems** 31:065003 (16p), 2015. 2015, 31, 065003, 16

uniqueness for F_0 only under strong additional assumptions:

- symmetry
- analyticity close to boundary, C^4 elsewhere

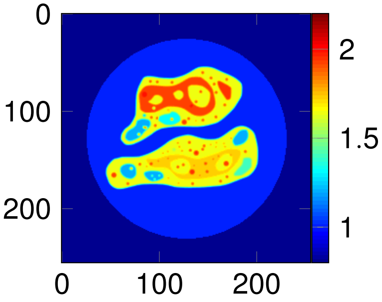
M. V. Klibanov. *On the recovery of a 2-D function from the modulus of its Fourier transform.* **J. Math. Anal. Appl.** 323:818–843, 2006.

exact diffraction pattern and photon counts



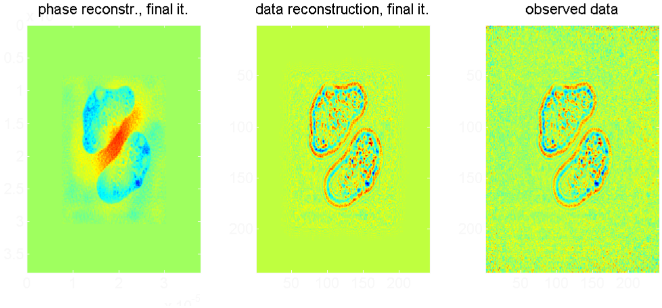
Expected total number of photon counts $= 10^6$

simulated phase object and reconstruction



t	10^3	10^4	10^5	10^6	10^7	10^8
L^2-fidelity	58.8	50.7	31.5	16.6	9.46	9.21
$\mathcal{S}$-fidelity	53.2	39.2	29.3	13.8	8.77	7.38

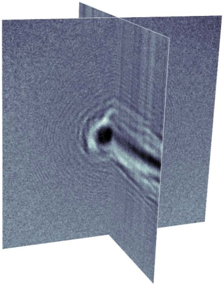
reconstruction of a cell from holographic experimental data in the Fresnel regime



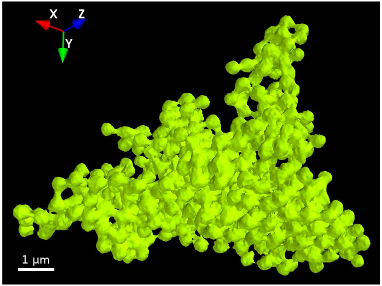
experimental data published in:

K. Giewekemeyer, S.P. Krüger, S. Kalbfleisch, M. Bartels, C. Beta, T. Salditt. *X-ray propagation microscopy of biological cells using waveguides as a quasipoint source.* **Phys. Rev. A** 83:023804. 2011

3D reconstructions from tomographic experimental data



$$\xrightarrow[\text{(by IRNM)}]{F_{\text{tomo}}^{-1}}$$



source:

S.Maretzke *Regularized Newton methods for simultaneous Radon inversion and phase retrieval in phase contrast tomography.* Master thesis. arXiv:1502.05073, 2015.

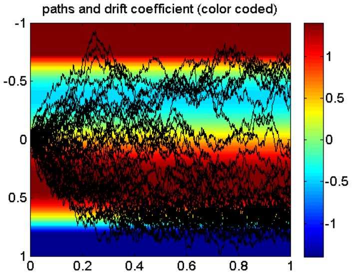
Outline

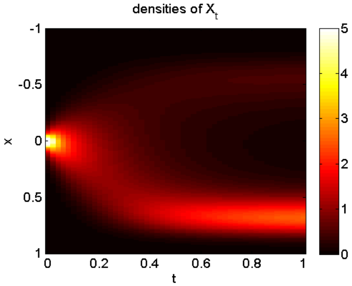
statement of the problem:

Consider a stochastic differential equation

$$d\mathbf{Y}_t = \mu(\mathbf{Y}_t)\,dt + \sigma(\mathbf{Y}_t)d\mathbf{W}_t\,.$$

inverse problem: Given the values $\mathbf{Y}_T^{(j)}$, $j = 1, \ldots, n$ of N independent paths starting at $\mathbf{Y}_0 = 0$ for some time $T > 0$ and given σ, estimate the drift coefficient μ!





Fokker-Planck equation

Assume that $\mathbf{Y}_t$ has a density $g(\cdot, t)$ w.r.t. the Lebesgue measure for all $t \in [0, T]$. Then g solves the Fokker-Planck equation (also called Kolmogoroff forward equation)

$$\frac{\partial g}{\partial t} = \mathrm{div}\left(\frac{\sigma\sigma^\top}{2}\,\mathrm{grad}\, g - \mu g\right).$$

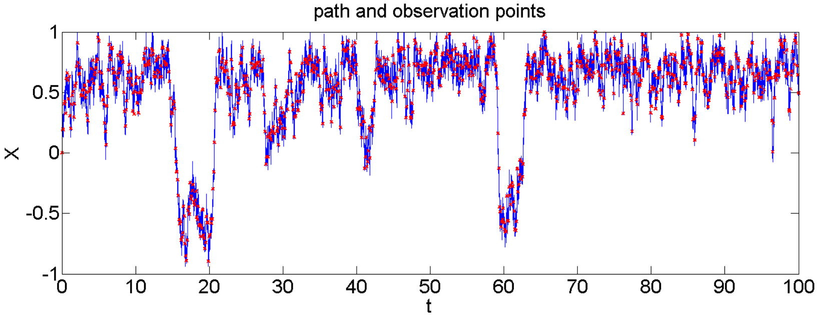
formulation as an operator equation: Introduce $F : H^s \to L^1$ with $s > d/2 + 1$ by $F(\mu) := g(\cdot, T)$ where g solves the Fokker-Planck equation with initial values $g(\cdot, 0) = \delta_0$.

Data consist of N independent samples $\mathbf{Y}_T^{(1)}, \ldots, \mathbf{Y}_T^{(N)}$ drawn from the distribution with density $g^\dagger = F(\mu^\dagger)$.

second scenario: equidistant observations of an ergodic process

A similar problem arises if a single path is observed at equidistant time points and ergodicity is assumed. The density g of the observations satisfies the stationary Fokker Planck equation

$$0 = \operatorname{div}\left(\frac{\sigma\sigma^\top}{2}\operatorname{grad} g - \mu g\right), \qquad \int g\,dx = 1.$$



negative log-likelihood

Let $Y_N := \frac{1}{N}\sum_{j=1}^N \delta_{\mathbf{Y}_j}$ denote the empirical measure.

negative log-likelihood:

$$\mathcal{S}_{Y_N}(g) = -\frac{1}{N}\ln \mathbb{P}_g[\mathbf{Y}_1, \dots, \mathbf{Y}_N] = -\frac{1}{N}\ln \prod_{j=1}^N g(\mathbf{Y}_j)$$
$$= -\frac{1}{N}\sum_{j=1}^N \ln g(\mathbf{Y}_j) = -\int \ln(g)\, dY_N.$$

nonnegative deterministic data-fidelity term:

$$\mathcal{T}_{g^\dagger}(g) = \mathbb{E}\left[\mathcal{S}_{Y_N}(g) - \mathcal{S}_{Y_N}\left(g^\dagger\right)\right] = \mathrm{KL}\left(g^\dagger; g\right)$$

where $\mathrm{KL}\left(g^\dagger; g\right) := \int g^\dagger \ln \frac{g^\dagger}{g}, dx$.

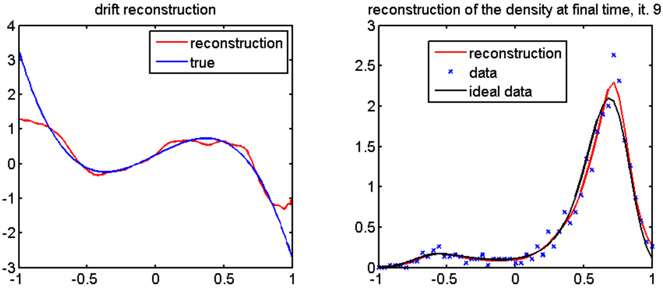
convergence in expectation

For the second scenario all assumptions of our convergence theorems both for generalized Tikhonov regularization and iteratively regularized Newton methods could be verified. The final result for an a-priori choice of α or the stopping index is convergence in expectation of the form

$$\mathbb{E}\left[\|\widehat{\boldsymbol{\mu}}_\alpha - \boldsymbol{\mu}^\dagger\|_{H^s}^2\right] \leq C\psi\left(\frac{1}{\sqrt{N}}\right), \qquad N \to \infty.$$

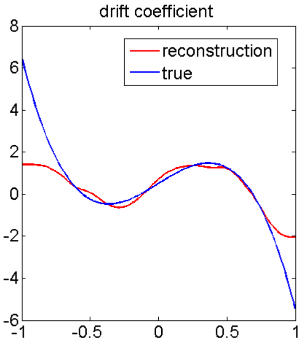
F. Dunker, T. Hohage. *On parameter identification in stochastic differential equations by penalized maximum likelihood.* **Inverse Problems** 30:095001, 2014.

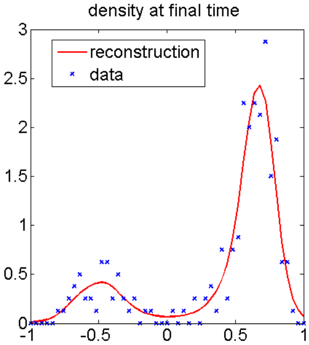
reconstructions from single path observations



N	125	250	500	1000
L^2-fidelity	0.28	0.22	0.18	0.14
$\mathcal{S}$-fidelity	0.18	0.14	0.11	0.096

reconstructions from many paths at fixed time



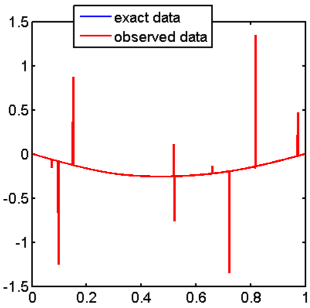


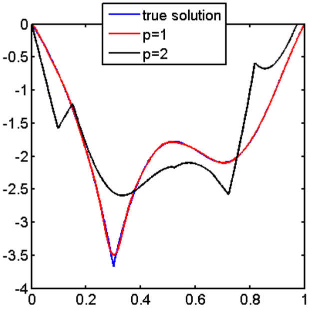
Outline

comparison of L^2 and L^1 data fidelity for impulsive noise

F = linear integral operator (two times smoothing)

$$\widehat{\boldsymbol{\mu}}_\alpha = \operatorname{argmin}_{\boldsymbol{\mu}\in L^2} \left[\|F\boldsymbol{\mu} - g^{\text{obs}}\|_{L^p}^p + \alpha\|\boldsymbol{\mu}\|_{L^2}^2 \right], \qquad p = 1, 2$$





Computation of L^1 minimizer via dual formulation, see

C. Clason, B. Jin, K. Kunisch. *A semismooth Newton method for L^1 data fitting with automatic choice of regularization parameters and noise calibration.* **SIAM J. Imaging Sci.** 3:199–231, 2010.

obstacle scattering

- aim: find boundary $\partial D = \{\boldsymbol{\mu}(\hat{x})\,\hat{x} \mid \hat{x} \in S^1\}$
- $v(x) = \exp(ikd \cdot x) + v_{\mathrm{s}}(x)$

$$\Delta v + k^2 v = 0 \text{ in } \mathbb{R}^2 \setminus \overline{D}$$

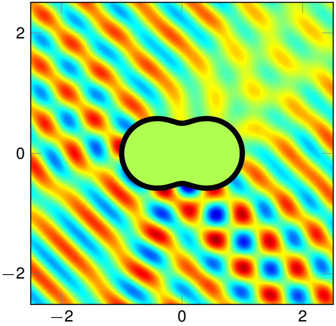
$$\frac{\partial v}{\partial n} = 0 \text{ on } \partial D$$

$$\sqrt{r}\left(\frac{\partial v_{\mathrm{s}}}{\partial r} - ikv_{\mathrm{s}}\right) \overset{r\to\infty}{\to} 0$$

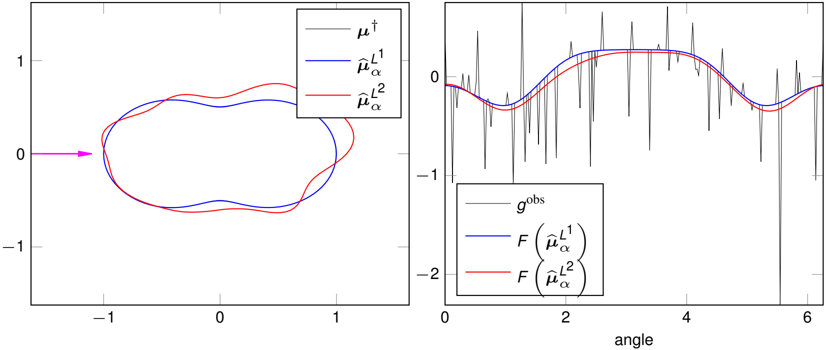
- $v(x) \approx \frac{e^{ik|x|}}{\sqrt{|x|}} v_\infty\left(\frac{x}{|x|}\right).$
- forward operator

$$F : H^s\left(S^1\right) \to L^1(S^1),$$

$$\boldsymbol{\mu}^\dagger \mapsto v_\infty$$



inverse obstacle scattering with impulsive noise



a limiting case

Let $\mathcal{Y} = L^1(\Omega)$ and

$$g^{\mathrm{obs}} = g^\dagger + \sum_{j=1}^{J} c_j \delta_{x_j} .$$

Recall that $L^1(\Omega)$ is isometrically embedded in $\mathcal{M}(\Omega)$, the space of signed finite Borel measures.

- $\|g^{\mathrm{obs}} - g^\dagger\|_{\mathcal{M}(\Omega)} = \sum_{j=1}^{J} |c_j|$, so the classical noise level may be large.
- Choose $\mathcal{S}_{g_1}(g_2) := \|g_1 - g_2\|_{\mathcal{M}}$, $\mathcal{T}_{g_1}(g_2) := \|g_1 - g_2\|_{L^1}$.
- $\mathcal{S}_{g^{\mathrm{obs}}}(g) = \|g - g^\dagger\|_{L^1} + \sum_{j=1}^{J} |c_j|$
- Hence, $\mathcal{S}_{g^{\mathrm{obs}}}(g) - \mathcal{S}_{g^{\mathrm{obs}}}(g^\dagger) = \mathcal{T}_{g^\dagger}(g)$, so for $C_{\mathrm{err}} = 1$ we get

$$\mathrm{err} = 0.$$

T. Hohage, F. Werner. *Convergence rates for inverse problems with impulsive noise*. **SIAM J. Numer. Anal.** 52:1203–1221, 2014.

impulsive noise model and improved error bound

There exist $\eta, \epsilon \geq 0$ and a measurable $\mathbb{M}_\eta \subset \mathbb{M}$ such that

$$|\mathbb{M}_\eta| \leq \eta, \qquad \|g^\dagger - g^{\mathrm{obs}}\|_{L^1(\mathbb{M}\setminus\mathbb{M}_\eta)} \leq \epsilon\,.$$

assumptions:

- $\mathcal{X}$ Hilbert space and $F : \mathcal{X} \to W^{k,2}(\mathbb{M})$ Lipschitz continous bounded with $\mathbb{M} \subset \mathbb{R}^d$ and $k > d/2$
- variational source condition with $\psi(t) = ct^\mu$

Then

$$\|\widehat{\boldsymbol{\mu}}_\alpha - \boldsymbol{\mu}^\dagger\|_{\mathcal{X}} = \mathcal{O}\left(\epsilon^{\mu/2} + \eta^{\frac{\mu}{2+\mu}\left(\frac{k}{d}+\frac{1}{2}\right)}\right).$$

impulsive noise for infinitely smoothing operators

Recall noise model: $\exists \eta, \epsilon \geq 0$, $\mathbb{M}_\eta \subset \mathbb{M}$ measurable such that

$$|\mathbb{M}_\eta| \leq \eta, \qquad \|g^\dagger - g^{\text{obs}}\|_{L^1(\mathbb{M}\setminus\mathbb{M}_\eta)} \leq \epsilon\,.$$

For inverse problems in PDEs the forward operator F is very often not only finitely, but infinitely smoothing.

assumptions:

- F maps boundedly into a space of analytic functions
- logarithmic source condition

Then

- only logarithmic convergence rates in ϵ, but still high polynomial rates in η

T. Hohage, C. König, F. Werner. *Convergence Rates for Exponentially Ill-Posed Inverse Problems with Impulsive Noise*. arXiv: 1506.02126, 2015.