Convergence rates for variational regularization of inverse problems with Poisson and empirical process data

Yusufu Simayi, Benjamin Sprung and Thorsten Hohage

Institute for Numerical and Applied Mathematics, University of Göttingen, Lotzestr. 16-18, 37083, Göttingen, Germany

E-mail: y.simayi@math.uni-goettingen.de and hohage@uni-goettingen.de

2 October 2019

Abstract. We discuss variational regularization method for linear and nonlinear ill-posed inverse problems described by operator equations $F(f^\dagger) = g^\dagger$, where $g^\dagger$ is an integrable, non-negative function and the measured data are described either by a Poisson process or an empirical process with density $g^\dagger$. Such inverse problems occur for example in photonic imaging and parameter identification problems in stochastic differential equations, and variational regularization methods for such problems have been studied intensively in recent years. However, known rates of convergence for mildly ill-posed problems with such data are not optimal so far. For both noise models, Kullback-Leibler data fidelity terms appear naturally as negative log-likelihood functionals. We improve the reconstruction error bound in expectation for the mildly ill-posed case. In order to improve rates of convergence for variational regularization, the key tool of our analysis are deviation inequalities for Poisson and empirical process in negative Besov norms and interpolation inequalities in Besov spaces.

Keywords: Optimal order, source conditions, variational regularization, convergence rate, Poisson process, empirical process, Kullback-Leibler divergence.

1. Introduction

In this paper, we study a statistical convergence analysis for variational regularization method of linear and nonlinear ill-posed inverse problems with Poisson and empirical process data in Hilbert spaces. Suppose that the unknown quantity $f^\dagger$ is an element of a Hilbert space $\mathbb{X}$ and $g^\dagger$ is a non-negative, integrable function on the manifold $\mathbb{M}$. Then we consider an operator equation for a continuous (possibly nonlinear) operator $F : \mathbb{X} \to \mathbb{Y} \subset L^2(\mathbb{M})$ between the Hilbert spaces $\mathscr{B} \subset \mathbb{X}$ and $\mathbb{Y}$ by

$$F(f^\dagger) = g^\dagger. \tag{1}$$

For the case of inverse problems with Poisson data, the ideal data $g^\dagger \in L^1(\mathbb{M})$ can be interpreted as a photon density on some measurement manifold $\mathbb{M} \subset \mathbb{R}^d$ and the observed data will be drawn from a Poisson point process with density $ng^\dagger$, where the parameter $n > 0$ is often interpreted as an exposure time. In the case of inverse problems with data given by empirical processes, $g^\dagger$ corresponds to probability density of the i.i.d random variables and the data will be sampled from i.i.d random variables with probability density $g^\dagger$.

Inverse problems with Poisson data have been intensively studied over the last two decades since they have many applications in photonic imaging, medicine, astronomy [4], coherent x-ray imaging [16], fluorescence microscopy and positron emission tomography [5, 33]. In these inverse problems, the photon count data are Poisson distributed for some physical reasons. There exists various efficient methods applied for ill-posed linear and nonlinear inverse problems with Poisson data, see, e.g., [2, 3, 4, 5, 18, 16, 23, 37]. Inverse problems with empirical data described by stochastic differential equations also studied by Dunker and Hohage [6] and obtained convergence rates for penalized maximum likelihood estimator.

In order to construct a stable approximation to $f^\dagger$, we consider a Tikhonov type estimator

$$\widehat{f}_\alpha \in \operatorname*{argmin}_{f \in \mathscr{B}} \left[\mathscr{S}(G_n, F(f)) + \frac{\alpha}{2} \|f\|_{\mathbb{X}}^2 \right], \tag{2}$$

where $\alpha > 0$ is a regularization parameter, G_n describes the observed data, $\mathscr{S}$ is the non-quadratic data fidelity term and $\mathscr{R}(f) = \frac{1}{2}\|f\|^2$ is quadratic penalized term. If $\mathscr{S}$ is also quadratic with Hilbert space, then (2) is a usual Tikhonov regularization. If the operator F is linear and the data fidelity and penalized terms are convex in its second argument, then the minimization problem (2) is convex.

The Kullback-Leibler divergence has been studied as a data fidelity term for inverse problems with Poisson data in a number of literature [2, 3, 5, 6, 16, 18, 32, 37] and for the inverse problem with empirical process data by Dunker and Hohage [6] under deterministic and stochastic error assumptions. Moreover, it also corresponds to other types of noise models, see, e.g., [32]. Dunker and Hohage [6] even show that Kullback- Leibler type data fidelity term provide better results than the quadratic data fidelity term by using Monte-Carlo simulation.

There exist a large number of literature on regularization methods for linear and nonlinear ill-posed inverse problems with Poisson and empirical process data. In particular, Nowak et al. [24] proposed a Bayesian multiscale framework for Poisson inverse problems with discrete case. Szkutnik [27] study a quasi-maximum likelihood approach for projection type estimator. Reynaud-Bouret [25], Antoniadis and Bigot [2] and Cavalier et al [5] achieved optimal convergence rates of linear ill-posed problems with Poisson data by. Recently, Werner and Hohage [16, 18, 37] intensively studied convergence rates for variational regularization methods of nonlinear inverse problems in Banach space with Poisson data in the context of discrete and continuous setting and further show the efficient algorithms for such inverse problems with several applications over the last five years. Furthermore, Dunker and Hohage [6] also studied convergence rates for variational regularization with empirical process data and derived general convergence results of the risk for penalized maximum likelihood estimator with Tikhonov-type and Newton-type regularization methods. Although Werner and Hohage [37], Dunker and Hohage [6] obtained convergence rates for variational regularization of inverse problems with Poisson and empirical process data in terms of an index function, respectively, these results cannot show order optimality for mildly ill-posed inverse problems. Moreover, Antoniadis and Bigot [2] achieved optimal convergence rates for Poisson inverse problem by using a Wavelet-Galerkin approximation method in certain function spaces, however their approach is restricted to linear operators. Apart from those results, Weidling et al [36] in recent years developed a method for verification of variational source conditions in Hilbert and Banach spaces under some smoothness assumptions. These results can accelerate our convergence analysis for the estimator (2) to achieve better reconstruction error bounds in a stochastic setting.

In this paper, we will further improve convergence rates with explicit exponent for variational regularization of mildly ill-posed nonlinear inverse problems with Poisson and empirical process data simultaneously. To derive better reconstruction error bounds, we will first need to establish deviation inequalities for Poisson and empirical processes in negative Besov norms and apply the interpolation theory in terms of the Besov norms. With the help of these results we are able to improve the convergence rates. Moreover, to derive order optimal rates for variational regularization, we will state a conjecture to bound noise processes in terms of the negative Besov norms with the limiting case.

The structure of this paper is organized as follows: In section 2, we give some properties of Poisson and empirical processes and recall recent results on uniform concentration inequalities for these processes obtained by Reynaud-Bouret [25] and Massart [21], respectively. In section 3 we will obtain large deviation inequalities for the suprema of Poisson and empirical processes in negative Besov norms based on results of uniform concentration inequalities for Poisson and empirical processes [21, 25] and we describe our conjecture. In section 4, we introduce source conditions in the form of variational inequality and present the previous convergence rate results derived by Hohage and Werner [18], which is helpful to clarify that how much small their convergence rates are. Finally, our main results will be shown in Theorem 4.10 and optimal orders of convergence rates in expectation are also stated under our conjecture.

2. Preliminaries for Poisson and empirical processes

Let $\mathbb{M} \subset \mathbb{R}^d$ be a Riemannian manifold, then a point process on $\mathbb{M}$ can be regarded as a random collection of points $\{x_1, \cdots, x_N\} \subset \mathbb{M}$, where the positions x_j of detected points and the total observed points N are random. For the given observation $X_j = x_j$ on $\mathbb{M}$, the sum of Dirac-measures can be defined by

$$G = \sum_{j=1}^{N} \delta_{x_j}, \quad \text{for } j = 1, 2 \cdots, N \tag{3}$$

at the point positions and

$$G(\mathbb{A}) = \#\{j = \{1, 2, \cdots, N\} | x_j \in \mathbb{A}\}$$

are the number of points in the measurable subset $\mathbb{A} \subset \mathbb{M}$. As described by Kingman [20], the Dirac measure (3) is called Poisson point process (or Poisson process) with intensity $g^\dagger \in L^2(\mathbb{M})$ if the following properties are satisfied:

(i) For any choice of disjoint and measurable subsets $\mathbb{A}_1, \cdots, \mathbb{A}_m \subset \mathbb{M}$, the random variables $G(\mathbb{A}_1), \cdots, G(\mathbb{A}_m)$ are stochastically independent.

(ii) For any measurable subset $\mathbb{A} \subset \mathbb{M}$, we have $\mathbf{E}[G(\mathbb{A})] = \int_{\mathbb{A}} g^\dagger dx$.

Note that if $g^\dagger$ is normalized as a density function for an exposure time $n > 0$, then $n = \mathbf{E}[N]$ is regarded as the expected total number of points.

One has shown that for each measurable subset $\mathbb{A} \in \mathbb{M}$ the random variable $G(\mathbb{A})$ is Poisson distributed with parameter $\lambda := \int_{\mathbb{A}} g^\dagger dx$, which is given by

$$\mathbf{P}[G(\mathbb{A}) = k] = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \text{for } k \in \mathbb{N},$$

see Theorem 1.11.8 in [20]. Furthermore, let $G = \sum_{j=1}^{N} \delta_{x_j}$ be a Poisson process with intensity $g^\dagger \in L^2(\mathbb{M})$, then the process G conditioned on $G(\mathbb{M}) = N$ is an empirical process with parameter N and probability measure $\xi(\mathbb{A})$ is give by

$$\xi(\mathbb{A}) := \frac{\int_{\mathbb{A}} g^\dagger dx}{\int_{\mathbb{M}} g^\dagger dx}$$

and the distribution

$$\mathbf{P}(G(\mathbb{A}_1) = n_1, \cdots, G(\mathbb{A}_m) = n_m | G(\mathbb{M}) = N) = N! \prod_{i=0}^{m} \frac{\xi(\mathbb{A}_t)^{n_i}}{n_i!},$$

see the proof of Proposition 2.2 [18].

Since the Poisson process $G = \sum_{j=1}^{N} \delta_{x_j}$ can be seen as a random measure, we can define integrals over continuous function $\Gamma : \mathbb{M} \to \mathbb{R}$ with respect to G. i.e., $\int_{\mathbb{M}} \Gamma dG = \sum_{j=1}^{N} \Gamma(x_j)$. For any complex continuous function $\Gamma : \Omega \to \mathbb{C}$ with respect to G, we have

$$\mathbf{E}\left[\int_{\mathbb{M}} \Gamma dG\right] = \int_M g^\dagger \Gamma dx, \quad \mathbf{Var}\left[\int_{\mathbb{M}} \Gamma dG\right] = \int_M g^\dagger |\Gamma|^2 dx \tag{4}$$

whenever the integral on the right hand sides exist, (see [20]).

As introduced by Werner and Hohage in [18], we need an additional parameter $n > 0$, which is called an exposure time. Assume the data $\widetilde{G}_n$ are drawn from a Poisson process with intensity $ng^\dagger$ and define a scaled Poisson process $G_n = \frac{1}{n}\widetilde{G}_n$, where the scaling factor $\frac{1}{n}$ ensure that the expectation of the integral function is independent of $n > 0$, see ([18]). Thus, (4) yields

$$\mathbf{E}\left[\int_{\mathbb{M}} g dG_n\right] = \int_M g^\dagger g dx, \quad \mathbf{Var}\left[\int_{\mathbb{M}} g dG_n\right] = \frac{1}{n} \int_M g^\dagger g^2 dx. \tag{5}$$

This indicate that the noise level should be determined by the scaling factor $\frac{1}{\sqrt{n}}$. We consider the convergence rates and optimality for the convergence rates of estimator to inverse problem (1) with the data G_n in the limit $n \to \infty$.

In addition, we also consider other types of stochastic process. An empirical process is a finite and infinite sequence of independent random variables and it occurs in non-parametric statistics. If the observable data are described by independent and identically distributed random variables $Y_1, \cdots, Y_n$ and each of which has probability density $g^\dagger$. This problem can be formulated as a nonlinear ill-posed operator equation as (1). For the given observations $Y_j = y_j$, the empirical measure is given by

$$G_n = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}, \tag{6}$$

where δ_{Y_j} are the Direc measure. For any Borel measurable subset $\mathbb{A} \subset \mathbb{M}$, we have

$$G_n = \frac{1}{n} \sum_{j=1}^n \mathbb{I}_\mathbb{A}(Y_j) = \frac{\#\{j \le n : Y_j \in \mathbb{A}\}}{n}.$$

Analogously, the empirical process $G_n = \frac{1}{n} \sum_{j=1}^n \delta_{y_j}$ also the random measure such that the integrals over complex continuous function $\tilde{\Gamma} : \mathbb{M} \to \mathbb{R}$ w.r.t. G_n can be defined as

$$\int_\mathbb{M} \tilde{\Gamma} \mathrm{d}G_n = \sum_{j=1}^N \tilde{\Gamma}(x_j).$$

For any complex continuous function $\tilde{\Gamma} : \mathbb{M} \to \mathbb{C}$ w.r.t. G_n, we have

$$\mathbf{E}\left[\int_\mathbb{M} \tilde{\Gamma} \mathrm{d}G_n\right] = \int_M g^\dagger \tilde{\Gamma} \mathrm{d}x, \quad \mathbf{Var}\left[\int_\mathbb{M} \tilde{\Gamma} \mathrm{d}G_n\right] = \frac{1}{n} \int_M g^\dagger |\tilde{\Gamma}|^2 \mathrm{d}x \tag{7}$$

whenever the integral on the right hand sides exist.

Finally, we will recall uniform concentration inequalities for the suprema of Poisson process and empirical process, respectively. More precisely, the uniform concentration inequality for the suprema of empirical processes has been derived by Massart in Theorem 3 [21]. Similarly, such kind of concentration inequality for the suprema of Poisson process has been obtained by Reynaud-Bouret in Corollary 2 [25] . These concentration inequalities are actually established based on the Talagrand's concentration inequalities in product spaces [28].

Lemma 2.1. . *Let G_n be either a Poisson process or empirical process with intensity $g^\dagger$ and let $\mathscr{F} \subset L^\infty(\mathbb{M})$ be a countable family of functions with $\|\varphi\|_\infty \le b$ for all $\varphi \in \mathscr{F}$. Moreover, let*

$$Z := n \sup_{\varphi \in \mathscr{F}} \left| \int_\mathbb{M} \varphi(\mathrm{d}G_n - g^\dagger \mathrm{d}x) \right| \quad and \quad v := n \sup_{\varphi \in \mathscr{F}} \int_\mathbb{M} \varphi^2 g^\dagger \mathrm{d}x. \tag{8}$$

(i) If G_n is a Poisson process, then the following inequality holds true

$$\mathbf{P}\left[Z \ge (1+\varepsilon)\mathbf{E}[Z] + 2\sqrt{3v\eta} + \mu_1(\varepsilon)b\eta\right] \le \exp(-\eta) \tag{9}$$

for $\eta, \varepsilon > 0$, where $\mu_1(\varepsilon) = \frac{5}{4} + \frac{32}{\varepsilon}$.
(ii) If G_n is an empirical process, then the following inequality holds true

$$\mathbf{P}\left[Z \ge (1+\varepsilon)\mathbf{E}[Z] + 2\sqrt{2v\eta} + \mu_2(\varepsilon)b\eta\right] \le \exp(-\eta) \tag{10}$$

for $\eta, \varepsilon > 0$, where $\mu_2(\varepsilon) = \frac{5}{2} + \frac{32}{\varepsilon}$.

Actually, Lemma 2.1 can be seen as a combination of the results of Corollary 2 in [25] and Theorem 3 in [28]. In recent years, Werner and Hohage [37] and Dunker and Hohage [6] derived the deviation inequalities for the noise process on the Sobolev ball based on the results of Lemma 2.1, respectively. However, these deviation inequalities do not yield optimal bounds for the noise error. For this reason, we further study deviation inequalities for these processes in terms of negative Besov norms.

3. Deviation inequalities for Poisson and empirical processes in negative Besov norms

The main purpose of this section is to derive deviation inequalities for Poisson and empirical processes in negative Besov norms with Fourier analysis, which can be established based on the results obtained by Reynaud-Bouret [25] and Massart [21], respectively. Before we establish our deviation inequalities, we will recall some basic properties of the periodic Besov spaces

3.1. Periodic distributions and their Fourier transform

Let $\mathbb{T}^d := \mathbb{R}^d/\mathbb{Z}^d$ denote the d-dimensional torus and let $\mathscr{D}(\mathbb{T}^d)$ denote the space of all complex-valued, infinitely differentiable, multi-periodic functions with periodicity 2π. Equipped with the topology generated by the family of semi-norms

$$\|P\|_r := \sup_{x \in \mathbb{T}^d} |\partial^r P(x)| \quad \text{with a multi-index } r = (r_1, \cdots, r_d) \in \mathbb{N}_0^d,$$

the space $\mathscr{D}(\mathbb{T}^d)$ becomes a locally convex topological vector space. The topological dual space $\mathscr{D}'(\mathbb{T}^d)$ of $\mathscr{D}(\mathbb{T}^d)$ is the space of periodic distribution. We use the common notation $\langle P, \varphi \rangle := P(\varphi)$ for $\in \mathscr{D}'(\mathbb{T}^d)$ and $\varphi \in \mathscr{D}(\mathbb{T}^d)$. It follows from the definition of the topology of $(D)(\mathbb{T}^d)$ that for any $P \in \mathscr{D}'(\mathbb{T}^d)$ there exist constants $M \in \mathbb{N}$ and $c > 0$ such that

$$|\langle P, \varphi \rangle| \le c \sum_{|r| \le M} \|\varphi\|_r \qquad \text{for all } \varphi \in \mathscr{D}(\mathbb{T}^d). \tag{11}$$

As the trigonometric monomials $e_\kappa(x) := (2\pi)^{-d/2} e^{i\kappa \cdot x}$, $\kappa \in \mathbb{Z}^d$ obviously belong to $\mathscr{D}(\mathbb{T}^d)$, the Fourier transform $\widehat{P} : \mathbb{Z}^d \to \mathbb{C}$ of any periodic distribution $P \in \mathscr{D}'(\mathbb{T}^d)$ is well defined by

$$\widehat{P}(\kappa) = \langle P, e_\kappa \rangle, \quad \kappa \in \mathbb{Z}^d.$$

It follows from (11) that $|\widehat{P}(\kappa)| \le C_M (1 + |\kappa|)^{-M}$ for some $C_M > 0$, and this can be used to show that P can be represented by its Fourier series

$$P = \sum_{\kappa \in \mathbb{Z}^d} \widehat{P}(\kappa) e_\kappa$$

with convergence in $\mathscr{D}'(\mathbb{T}^d)$. For more details, we refer to [11, 13].

3.2. Definition of periodic Besov spaces

Let $\phi_0(x)$ be the characteristic function of the unit box in $\mathbb{R}^d$. We will need a smooth ϕ_0 and $\phi_l(x)$ be a system of the corresponding dyadic resolution of unity, which are given by

$$\phi_0(x) := 1, \quad \text{for } |x|_\infty \le \frac{1}{2} \quad \text{and} \quad \phi_0(x) := 0 \quad \text{for else}$$

and

$$\phi_l(x) := \phi_0(2^{-l}x) - \phi_0(2^{-l+1}x), \quad \text{for } l \in \mathbb{N}.$$

Note that this defines a partition of unity, i.e., $\sum_{l=0}^{\infty} \phi_l(x) = 1$ for $x \in \mathbb{R}^d$.

For $P \in \mathscr{D}'(\mathbb{T}^d)$ and $l \in \mathbb{N}_0$, we define

$$P_l := \sum_{\kappa \in \mathbb{Z}^d} \phi_l(\kappa) \widehat{P}(\kappa) e_\kappa. \tag{12}$$

As the sum is finite, we have $P_l \in \mathscr{D}(\mathbb{T}^d)$. Moreover, $P = \sum_{l=0}^{\infty} P_l$ with convergence in $\mathscr{D}'(\mathbb{T}^d)$.

Let $s \in \mathbb{R}$ and $p, q \in [1, \infty]$ be given parameters of smoothness, integrability and summability respectively. Then we define the periodic Besov space as

$$B_{p,q}^s(\mathbb{T}^d) := \{P \in \mathscr{D}'(\mathbb{T}^d) : \|P\|_{B_{p,q}^s(\mathbb{T}^d)} < \infty\},$$

where the Besov norm is given by

$$\|P\|_{B_{p,q}^s(\mathbb{T}^d)} := \begin{cases} \left(\sum_{l \geq 0} 2^{slq} \|P_l\|_{L^p(\mathbb{T}^d)}^q\right)^{1/q}, & q < \infty, \\ \sup_{l \geq 0} 2^{sl} \|P_l\|_{L^p(\mathbb{T}^d)}, & q = \infty. \end{cases}$$

The Besov spaces $B_{p,q}^s(\mathbb{T}^d)$ are Banach spaces such that

$$\mathscr{D}(\mathbb{T}^d) \hookrightarrow B_{p,q}^s(\mathbb{T}^d) \hookrightarrow \mathscr{D}'(\mathbb{T}^d),$$

where the first embedding is dense if p, q are finite. In particular we have $B_{2,2}^s(\mathbb{T}^d) = H_2^s(\mathbb{T}^d)$ for $p = q = 2$ and for all $s \in \mathbb{R}$, where H_2^s denotes a Sobolev space. If the smoothness $s = 0$, then we have $B_{2,2}^0(\mathbb{T}^d) = L^2(\mathbb{T}^d)$.

3.3. Bounds on Besov norms of empirical and Poisson noise.

Since we are interested in the Besov space $B_{p,q}^s$ with a general bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$ and study deviation inequalities for Poisson and empirical processes on this domain, we define a Besov space with the norm by

$$\|P\|_{B_{p,q}^s(\mathbb{M})} := \inf\left\{\|\tilde{P}\|_{B_{p,q}^s} : \tilde{P} \in B_{p,q}^s(\mathbb{R}^d), \ \tilde{P}|_{\mathbb{M}} = P \text{ in the sense of}\mathscr{D}'(\mathbb{M})\right\},$$

where the infimum is taken over all extensions $\tilde{P} \in B_{p,q}^s(\mathbb{R}^d)$. In the following, we study $P = G_n - g^\dagger$ where G_n is either an empirical process of a rescaled Poisson process with density $g^\dagger$. We want to bound

$$Z^{(s)} := \|P\|_{B_{2,\infty}^{-s}(\mathbb{M})} = \sup_{l \geq 0} Z_l^{(s)} \qquad \text{where} \qquad Z_l^{(s)} := 2^{-sl} \|P_l\|_{L^2(\mathbb{M})}.$$

To apply Lemma 2.1 to each P_l, we introduce

$$\mathscr{S}_l^s := \{\mathfrak{f} \in L^2(\mathbb{M}) : \|\mathfrak{f}\|_{L^2} \leq 2^{-sl} \text{ and } \widehat{\mathfrak{f}}(\kappa) = 0 \text{ if } \kappa \notin \text{supp } \phi_l\}.$$

This gives that

$$Z_l^{(s)} = \sup_{\{\mathfrak{f} \in L^2(\mathbb{M}) : \|\mathfrak{f}\|_{L^2} \leq 2^{-sl}\}} \left|\int_{\mathbb{M}} P_l(x)\mathfrak{f}(x)\mathrm{d}x\right|$$

$$= \sup_{\|\mathfrak{f}\|_{L^2(\mathbb{M})} \leq 2^{-sl}} \left|\sum_{\kappa \in \mathbb{Z}^d} \widehat{\mathfrak{f}}(\kappa)\widehat{P_l}(\kappa)\right| = \sup_{\|\mathfrak{f}\|_{L^2(\mathbb{M})} \leq 2^{-sl}} \left|\sum_{\kappa \in \mathbb{Z}^d} \widehat{P}(\kappa)\phi_l(\kappa)\widehat{\mathfrak{f}}(\kappa)\right|$$

$$\leq \sup_{\|\mathfrak{f}\|_{L^2(\mathbb{M})} \leq 2^{-sl}} \left|\sum_{\phi_l(\kappa) \neq 0} \widehat{\mathfrak{f}}(\kappa)\widehat{P}(\kappa)\right|$$

$$\leq \sup_{\mathfrak{f} \in \mathscr{S}_l^s} \left|\int_{\mathbb{M}} \mathfrak{f}(x)(\mathrm{d}G_n - g^\dagger \mathrm{d}x)\right|. \tag{13}$$

In order to derive deviation inequalities, we need to bound the expectation of $Z^{(s)}$ in Besov spaces with the negative smoothness index $s > d/2$.

Lemma 3.1 (**Expectation of Poisson and empirical processes**). *Let G_n be either a rescaled Poisson process or an empirical process with intensity $g^\dagger \in L^\infty(\mathbb{M})$. Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded domain with the Lipschitz boundary $D > 0$. Assume that $Z^{(s)}$ is defined by () with the smoothness index $s > \frac{d}{2}$. Then there exists a constant C such that*

$$\mathbf{E}\left[\|G_n - g^\dagger\|_{B_{2,\infty}^{-s}(\mathbb{M})}\right] < \frac{CD}{\sqrt{n}}\|g^\dagger\|_{L^1(\mathbb{M})}. \tag{14}$$

Proof. Using the continuous embedding results in Besov spaces we know that $B_{2,2}^{-s} \subset B_{2,\infty}^{-s}$ for $s > d/2$, then

$$\left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}(\mathbb{M})} \leq C_1 \left\| G_n - g^\dagger \right\|_{B_{2,2}^{-s}(\mathbb{M})} \tag{15}$$

holds true for some constant $C_1 > 0$. From Lemma A.3 in Werner and Hohage [37], we have

$$\mathbf{E}\left[\left\| G_n - g^\dagger \right\|_{H^{-s}(\mathbb{M})} \right] \leq \frac{C_2 D}{\sqrt{n}} \| g^\dagger \|_{L^1(\mathbb{M})}$$

for some constant $C_2 > 0$. Since $H^{-s}(\mathbb{M}) = B_{2,2}^{-s}(\mathbb{M})$ for $s > d/2$, by taking expectation on both side of (15) we immediately derive (14) with $C = C_1 C_2 > 0$. $\qquad\square$

For the limiting case $s = d/2$ the expectation of the noise process $\| G_n - g^\dagger \|_{B_{2,\infty}^{-d/2}(\mathbb{M})}$ is not easy to bound. Therefore, we will state about the result as a conjecture.

Conjecture 3.2. *For an empirical or a Poisson process G_n with intensity $g^\dagger$ and the expected count n, we have*

$$\sup_{n \in \mathbb{N}} \mathbf{E}\left[\frac{\sqrt{n}}{(\ln n)^m} \| G_n - g^\dagger \|_{B_{2,\infty}^{-d/2}} \right] < \infty. \tag{16}$$

for $m \geq 0$.

As we seen in Lemma 3.1, the expectation for the noise process can be bounded in the norms of Besov space $B_{2,\infty}^s(\mathbb{M})$ with smoothness $s > d/2$. For the Gaussian white noise process, this conjecture holds true for $s = d/2$ and for any $p \in [1, \infty)$, which has been proven by Veraar [34] in Proposition 2.3. If one can prove this conjecture for Poisson and empirical processes in future, then it will give us an optimal rate as we obtain in Section 4. Now we are able to show the deviation inequalities for Poisson and empirical processes in negative Besov norms in the following theorem:

Theorem 3.3. *Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded d-dimensional Lipschitz manifold. Let $B_{2,\infty}^{-s}(\mathbb{M})$ denote the negative Besov space with index $s > \frac{d}{2}$.*

(i) If G_n be an empirical process with intensity $g^\dagger$ and constant $c_1 > 0$ such that

$$\mathbf{P}\left(\| G_n - g^\dagger \|_{B_{2,\infty}^{-s}} \geq \frac{\tilde{C} + \eta}{\sqrt{n}} \right) \leq \exp\left(-\frac{\eta}{c_1} \right) \tag{17}$$

holds true for all $n, \eta \geq 1$, where $c_1 = \sqrt{8 C_3} + 34.5$ with $C_3 > 0$ and $\tilde{C} = 2CD > 0$.

(ii) If G_n be a Poisson process with intensity $g^\dagger$ and constant $c_2 > 0$ such that

$$\mathbf{P}\left(\| G_n - g^\dagger \|_{B_{2,\infty}^{-s}} \geq \frac{\tilde{C} + \eta}{\sqrt{n}} \right) \leq \exp\left(-\frac{\eta}{c_2} \right) \tag{18}$$

holds true for all $n, \eta \geq 1$, where $c_2 = \sqrt{12 C_3} + 33.25$.

Proof. The proof of the theorem can be established based on the concentration inequalities given by Lemma 2.1, and we already bound the expectation of $Z^{(s)}$. Thus, it is sufficient to control the constants b and v respectively. The proof of these two processes are quite similar such that we only proof here for Poisson process. Since $Z^{(s)} := \sup_{l \geq 0} 2^{-sl} \| P_l \|_{L^2}$, we can control b by

$$b = \sup_{l \geq 0} \sup_{\mathfrak{f} \in \mathscr{S}_l^s} \| \mathfrak{f} \|_\infty \leq \sup_{l \geq 0} \sup_{\mathfrak{f} \in \mathscr{S}_l^s} 2^{\frac{d}{2}l} \| \mathfrak{f} \|_{L^2}$$

$$\leq \sup_{l \geq 0} 2^{-(s - \frac{d}{2})l} \leq 1, \tag{19}$$

To control the constant v, we use Hölder's inequality that

$$v = n \sup_{l \geq 0} \sup_{\mathfrak{f} \in \mathscr{S}_l^s} \int_{\mathbb{M}} \mathfrak{f}^2(x) g^\dagger \, \mathrm{d}x \leq n \sup_{l \geq 0} \| \mathfrak{f}^2 \|_{L^\infty(\mathbb{M})} \| g^\dagger \|_{L^1(\mathbb{M})}$$

$$\leq C_3 n \sup_{l \geq 0} 2^{dl} \|\mathfrak{f}\|^2_{L^2(\mathbb{M})} \|g^\dagger\|_{L^1(\mathbb{M})}$$

$$\leq C_3 n \sup_{l \geq 0} 2^{dl} 2^{-2sl} \|g^\dagger\|_{L^1(\mathbb{M})}$$

$$\leq C_3 n \|g^\dagger\|_{L^1(\mathbb{M})}. \tag{20}$$

for some constant $C_3 > 0$. We apply (10) in Lemma 2.1 to $Z^{(s)}$:

$$\mathbf{P}\left[Z^{(s)} \geq (1+\varepsilon)\mathbf{E}[Z^{(s)}] + \frac{\sqrt{8\widehat{\eta}v}}{n} + \frac{\widehat{\eta}\mu_2(\varepsilon)b}{n}\right] \leq \exp(-\widehat{\eta}) \tag{21}$$

for all $\widehat{\eta} > 0$. By plugging (14), (30) and (31) into (32) and choosing $\varepsilon = 1$ and using $\mathbf{P}(A) = 1 - \mathbf{P}^c(A)$ and $\|g^\dagger\|_{L^1} = 1$, we obtain

$$\mathbf{P}\left[Z^{(s)} \leq \frac{2CD}{\sqrt{n}} + \frac{\sqrt{8C_3\widehat{\eta}}}{\sqrt{n}} + \frac{34.5}{n}\widehat{\eta}\right] \geq 1 - \exp(-\widehat{\eta})$$

for all $\widehat{\eta}, n > 0$. Since $\frac{1}{n} \leq \frac{1}{\sqrt{n}}$ and $\sqrt{\widehat{\eta}} \leq \widehat{\eta}$ for $n, \widehat{\eta} \geq 1$, we derive

$$\mathbf{P}\left[Z^{(s)} \leq \frac{2CD}{\sqrt{n}} + \left(\sqrt{8C_3} + 34.5\right)\frac{\widehat{\eta}}{\sqrt{n}}\right] \geq 1 - \exp(-\widehat{\eta})$$

for all $\widehat{\eta}, n > 0$. Setting $\eta = \left(\sqrt{8C_3} + 34.5\right)\widehat{\eta}$ and $\tilde{C} := 2CD$, we derive

$$\mathbf{P}\left(Z^{(s)} \geq \frac{\tilde{C} + \eta}{\sqrt{n}}\right) \leq \exp\left(-\frac{\eta}{\sqrt{8C_3} + 34.5}\right),$$

by taking $c_1 = \sqrt{8C_3} + 34.5$, we can derive the deviation inequality (28). As a consequence of the above argument, we can also establish a deviation inequality for Poisson process in negative Besov norms. $\quad\square$

Remark 3.4. *It is important to highlight that a deviation inequality for Gaussian white noise in negative Besov norm has been proven in Veraar in Corollary 3.7 [34]. They study regularity of the Gaussian white noise on the d-dimensional torus and it has paths in the Besov spaces $B^{-d/2}_{p,\infty}(\mathbb{T}^d)$ with $p \in [1,\infty)$. Unfortunately, we have not succeeded to derive the deviation inequalities for Poisson or empirical processes with negative smoothness index $s = \frac{d}{2}$ due to difficulties to bound $Z^{(s)}$.*

4. Convergence rates of inverse problem with Poisson and empirical data

4.1. Known convergence results for Poisson and empirical

In this section, we will recall the regularization properties of the variational methods. To estimate the smoothness of the unknown solution, we will use a well-known source condition, which can be written in the form of a variational inequality introduced by Hofmann et al. [14]. Since the derivation of convergence rates for inverse problems require a-priori knowledge about the unknown solution $f^\dagger$, we usually measure it by Bregman distance. The Bregman distance became an increasingly important tool to study a convergence analysis for variational methods of inverse problems.

Let $\mathscr{R} : \mathbb{X} \to (-\infty, \infty]$ is a proper convex functional. For a subgradient $f^* \in \partial\mathscr{R}(f^\dagger)$, the Bregman distance of $\mathscr{R}$ is given by

$$\mathscr{D}^{f^*}_\mathscr{R}(f, f^\dagger) := \mathscr{R}(f) - \mathscr{R}(f^\dagger) - \langle f^*, f - f^\dagger\rangle,$$

for all $f \in \mathbb{X}$, where $\langle\cdot, \cdot\rangle$ denotes the duality product of $\mathbb{X}^*$ and $\mathbb{X}$. It has been known that the Bregman distance is non-negative due to the convexity and $\mathscr{D}^{f^*}_\mathscr{R}(f^\dagger, f^\dagger) = 0$ if and only if $f = f^\dagger$. The Bregman distance depends on the penalty term $\mathscr{R}$ and the choice of f^*. If $\mathscr{R}(f) = \frac{1}{2}\|f - f_0\|^2_\mathbb{X}$ in Hilbert spaces, then we have $\mathscr{D}^{f^*}_\mathscr{R}(f, f^\dagger) = \frac{1}{2}\|f - f^\dagger\|^2_\mathbb{X}$. Throughout this paper we will restrict ourselves in Hilbert space setting. Hence, as in Flemming [8], we need the following variational source conditions:

Assumption 4.1 (Variational source condition). *Let $\mathscr{R} = \frac{1}{2}\|f\|_{\mathbb{X}}^2$. Assume that the variational inequality*

$$\frac{1}{4}\|f - f^\dagger\|_{\mathbb{X}}^2 \leq \frac{1}{2}\|f\|_{\mathbb{X}}^2 - \frac{1}{2}\|f^\dagger\|_{\mathbb{X}}^2 + \Phi\left(\mathscr{T}(g^\dagger, F(f))\right) \tag{22}$$

for all $f \in \mathbb{X}$ holds true, where Φ is concave monotonically increasing and $\Phi(0) = 0$.

The convergence results under some deterministic noise assumptions in Hilbert spaces has been studied by Flemming [9, 10], where they proved that the variational inequality (22) can be derived from $f^\dagger - f_0 = \Psi(F'[f^\dagger]^* F'[f^\dagger])\omega$ for some $\omega \in \mathbb{X}$ with the tangential condition

$$\|F'[f^\dagger](f - f^\dagger)\|_{\mathbb{Y}} \leq C\|F(f - f^\dagger)\|_{\mathbb{Y}}$$

for all $f \in \mathbb{X}$ and for constant $C > 0$, where $\Psi : [0, \infty) \to [0, \infty)$ is also index function and Ψ^2 is concave. $F : \mathscr{D}(F) \subset \mathbb{X} \to \mathbb{Y}$ is Fréchet differentiable in an open domain $\mathscr{D}(F)$. Furthermore, they also proved in [8] that spectral source conditions implies variational source conditions. Assumption 4.1 holds true for some index function $\Phi = \Phi_\Psi$ with $F = T$ and $\mathscr{T}(g^\dagger, g) = \|g - g^\dagger\|_{\mathbb{Y}}^2$, see [9, 10]. For instance,

$$\Psi(\lambda) = \lambda^\nu \quad \text{for some } \nu \in (0, 1/2) \quad \Rightarrow \quad \Phi(\lambda) = c_\lambda \lambda^{\frac{2\nu}{2\nu+1}} \quad \text{for } c_\lambda > 0$$

and

$$\Psi(\lambda) = (-\ln(\lambda))^{-p} \quad \text{for some } p > 0 \quad \Rightarrow \quad \Phi(\lambda) = (-\ln(\lambda))^{-2p} \quad \text{for } p > 0.$$

For more details, we refer to [9]. In Hilbert space setting, Assumption 4.1 is weaker than spectral source conditions. Weidling and Hohage [15] also verified the some advantages of variational source conditions over the spectral source conditions. Variational source conditions even can be applied for non-quadratic data fidelity and penalty functionals and they allow to use in Banach space setting etc.

As given by Grasmair [12], the bias (approximation error) can be bounded by the Fenchel conjugation of $-\Phi$, which is defined by

$$(-\Phi)^*(\gamma) = \sup_{\omega \geq 0}(\gamma\omega + \Phi(\omega)), \quad \gamma > 0. \tag{23}$$

It will be bounded alternatively by a function $\Psi : (0, \infty) \to [0, \infty]$, i.e., $\Psi(\gamma) = (-\Phi)^*(-\frac{1}{\gamma})$. For the detail of the Fenchel duality, we refer to [7]. The deterministic convergence rate results for ill-posed problems have been explored with different noise models, (see, e.g., [8, 9, 17, 18, 37]). Here we consider the minimization problems with Poisson and empirical data in Hilbert spaces for obtaining optimal convergence rates in terms of norm distance rather than the Bregman distance.

In Bayssian statistics, the negative log-likelihood functional always used as a data fidelity term if the measured data is random, (see, [18]). If G_n is a rescaled Poisson process, then the negative log-likelihood functional is given by

$$\mathscr{S}_0(G_n, g) = -\ln \mathbf{P}_g[G_n] = \int_{\mathbb{M}} g \mathrm{d}x - \int_{\mathbb{M}} \ln g \mathrm{d}G_n. \tag{24}$$

However, the negative log-likelihood for empirical process is slightly different from the data fidelity functional for Poisson process. For an empirical process G_n, the probability density is defined as $\mathbf{P}_g[y_1, \cdots, y_n] = \prod_{j=1}^n g(y_j)$. Then the corresponding negative log-likelihood is given by

$$\mathscr{S}_0(G_n, g) = -\frac{1}{n}\ln \mathbf{P}_g[y_1, \cdots, y_n] = -\int_{\mathbb{M}} \ln g \mathrm{d}G_n. \tag{25}$$

It is well-known that the the measured data will be accessed through the data fidelity functional. The minimum values of (24) and (25) has no importance and need to subtract $\mathscr{S}_0(G_n, g^\dagger)$. By taking expectation for the negative log-likelihood data fidelity terms (24) and (25) with the distance from $\mathscr{S}_0(G_n, g^\dagger)$, we have

$$\mathbf{E}\left[\mathscr{S}_0(G_n, g) - \mathscr{S}_0(G_n, g^\dagger)\right] = \int_{\mathbb{M}}\left[g - g^\dagger + g^\dagger \ln\left(\frac{g^\dagger}{g}\right)\right]\mathrm{d}x, \tag{26}$$

the right hand side of (26) is known as Kullback-Leibler divergence with $0\ln 0 = 0$ and $\ln(x) = \infty$ for $x \leq 0$ and $\mathrm{KL}(g, g^{\dagger})$ is a strictly convex and lower-semicontinuous functional on $\mathbf{L}^1(\mathbb{M})$. Note that Kullback-Leibler divergence $\mathrm{KL}(g, g^{\dagger}) \geq 0$ for all g and $\mathrm{KL}(g, g^{\dagger}) = 0$ if and only if $g = g^{\dagger}$. Now we can bound the error between the ideal and estimated data fidelity functionals by

$$
\begin{aligned}
\mathbf{err}(g) &:= \mathscr{S}_{\delta}(G_n, g) - \mathscr{S}_{\delta}(G_n, g^{\dagger}) - \mathrm{KL}(g, g^{\dagger}) \\
&= \begin{cases} \int_{\mathbb{M}} \ln\left(\dfrac{g}{g^{\dagger}}\right)(\mathrm{d}G_n - g^{\dagger}\mathrm{d}x) & \text{if } g \geq -\dfrac{\delta}{2} \ a.e. \\ \infty, & \text{else,} \end{cases}
\end{aligned}
\tag{27}
$$

which is called *effective noise level*, see [18]. As for error bounds of Gaussian and impulsive noise, we refer to [17, 36].

It can be seen from the error term (27) that the distance measures in $L^1(\mathbb{M})$ corresponds to the both negative log-likelihood described in (24) and (25). Since $|\ln(g/g^{\dagger})|$ is not uniformly bounded at the point 0 for some $f \in \mathscr{B}$ and the concentration inequality or deviation inequality can not be shown. Therefore, we recommend a modified parameter $\delta > 0$ for the data fidelity term $\mathscr{S}_{\delta}(G_n, g)$ such that

$$
\mathscr{S}_{\delta}(G_n, g) := \begin{cases} \int_{\mathbb{M}} g\mathrm{d}x - \int_{\mathbb{M}} \ln(g + \delta)(\mathrm{d}G_n + \delta\mathrm{d}x) & \text{if } g \geq -\dfrac{\delta}{2} \ a.e., \\ \infty, & \text{else} \end{cases}
\tag{28}
$$

and

$$
\mathscr{T}_{\delta}(g^{\dagger}, g) := \begin{cases} \mathrm{KL}(g^{\dagger} + \delta, g + \delta), & \text{if } g \geq -\dfrac{\delta}{2} \ a.e., \\ \infty, & \text{else.} \end{cases}
$$

Therefore, the error term (27) now can be written as

$$
\begin{aligned}
\mathbf{err}(g) &:= \mathscr{S}_{\delta}(G_n, g) - \mathscr{S}_{\delta}(G_n, g^{\dagger}) - \mathrm{KL}(g + \delta, g^{\dagger} + \delta) \\
&= \begin{cases} \int_{\mathbb{M}} \ln\left(\dfrac{g + \delta}{g^{\dagger} + \delta}\right)(\mathrm{d}G_n - g^{\dagger}\mathrm{d}x) & \text{if } g \geq -\dfrac{\delta}{2} \ a.e., \\ \infty, & \text{else.} \end{cases}
\end{aligned}
\tag{29}
$$

In general, the modified noise error (29) can be bounded with high probability by using a concentration inequality derived by Reynaud-Bouret [25] if we get $\|\ln\left(\frac{g+\delta}{g^{\dagger}+\delta}\right)\|_{\infty} < \infty$ under some proper assumptions. Hohage et al [6, 37] bounded the log-function in Sobolev ball and derived concentration inequalities for $\mathbf{err}(g)$. However, this convergence analysis cannot yield optimal rates, which we will state convergence results for variational regularization in the following theorem.

Theorem 4.2. *Suppose that Assumption 4.1 and (H6)-(H7) from [18] hold true with the $B_{2,1}^u(\mathbb{M})$-norm topology for $u > \frac{d}{2} > 0$ and if the Tikhonov functional (2) has a global minimizer $\widehat{f}_{\alpha}$, then we obtain*

$$
\left\|\widehat{f}_{\alpha} - f^{\dagger}\right\|_{\mathbb{X}}^2 \leq \frac{2\mathbf{err}(F(\widehat{f}_{\alpha}))}{\alpha} + 2\Psi(\alpha)
\tag{30}
$$

and

$$
\mathscr{T}_{\delta}(g^{\dagger}, F(\widehat{f}_{\alpha})) \leq 2\mathbf{err}(F(\widehat{f}_{\alpha})) + 2\alpha\Psi(2\alpha)
\tag{31}
$$

for all $\alpha > 0$. Moreover $F(\mathscr{B})$ is bounded with respect to $B_{2,2}^s(\mathbb{M})$-norm for $s > d/2$ if and only if $\alpha = \alpha(n)$ is chosen such that

$$
-\frac{1}{\alpha} \in \partial(-\Phi)\left(\frac{1}{\sqrt{n}}\right)
$$

as $n \to \infty$, we obtain the following convergence rate

$$
\mathbf{E}\left[\left\|\widehat{f}_{\alpha} - f^{\dagger}\right\|_{\mathbb{X}}^2\right] = O\left(\Phi\left(\frac{1}{\sqrt{n}}\right)\right)
\tag{32}
$$

as $n \to \infty$.

Theorem 4.2 can be seen as a particular case of Theorem 4.11 in Hohage and Werner [18] for Tikhonov-type regularization (2), where they derive convergence rates in terms of Bregman distance instead of quadratic distance norms. Unfortunately, these convergence results for variational regularization methods are not order optimal rates for Poisson and empirical process data. Due to the variational source condition, the convergence result in Theorem 4.2 that the smoothness of the true solution $f^\dagger$ is characterized by an index function Φ, in which case they could derive suboptimal rates in terms of the index function Φ and smoothing properties of the operator F. If $\mathbb{X}$ is a 2-convex Banach space and $\mathscr{R}(f) = \frac{1}{2}\|f\|_{\mathbb{X}}^2$, then it follows from the inequalities in Xu and Roach ([38]) or Sprung [26] that

$$\|f - f^\dagger\|_{\mathbb{X}} \leq C_{bd}\mathscr{D}(f, f^\dagger)^{\frac{1}{2}}$$

holds true for all $f \in \mathbb{X}$ with some constant $C_{bd} > 0$. This inequality also holds true for general penalty functional such as maximum entropy and sparsity functionals.

Recently, Weidling et al [36] showed for the verification of variational source conditions for the mildly ill-posed problems under some smoothness conditions, these approach are currently applied many other interesting inverse problems, e.g., [15, 19, 35]. Under this result, we can formulate the index function with an explicit power, see Theorem 4.1 in [36]. The following lemma will describe the specific formula of a variational source condition.

Lemma 4.3. *Let Assumption 4.1 is satisfied and $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau} \leq \rho$. Then there exists a constant $c > 0$ such that the solution $f^\dagger$ satisfies the variational source conditions (33) with*

$$\Phi(\alpha) = c\rho^{\frac{2u}{u+\tau}}\alpha^{\frac{\tau}{\tau+u}}.$$

Moreover, Ψ is given by

$$\Psi(\alpha) = c\rho^2\alpha^{\frac{\tau}{u}}. \tag{33}$$

Proof. The proof of the first argument follows from Theorem 4.1 in [36]. Now we need to calculate the Fenchel conjugate of Φ. By Fenchel formula, we have

$$\Psi(\gamma) = \sup_{\alpha > 0}\left[c\rho^{\frac{2u}{u+\tau}}\alpha^{\frac{\tau}{\tau+u}} - \frac{\alpha}{\gamma}\right]. \tag{34}$$

Since the first order $\Psi'(\bar{\alpha}) = 0$, we derive

$$\bar{\alpha} = \left(\frac{\tau + u}{u}\right)^{\frac{\tau+u}{\tau}}c^{-\frac{\tau+u}{\tau}}\rho^{-2}\gamma^{-\frac{\tau+u}{u}}.$$

Then by inserting this into (34) we obtain (33). □

Remark 4.4. *Under the assumptions of Lemma 4.3 and with parameter choice*

$$\alpha \sim (n\rho^4)^{-\frac{u}{2\tau+2u}},$$

we can write the convergence rate (33) more explicit form as follows

$$\mathbf{E}\left[\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2\right] = O\left(\rho^{\frac{4u}{2\tau+2u}}\left(\frac{1}{n}\right)^{\frac{\tau}{2\tau+2u}}\right), \tag{35}$$

as $n \to \infty$. However, the exponent $\frac{\tau}{2\tau+2u}$ is strictly smaller than the optimal rate with exponent $\frac{2\tau}{2\tau+2u+d}$ if and only if $d/2 < \tau + u$, which is always the case under our assumption $u > d/2$. Therefore, this order (35) is not satisfied the order optimality of convergence rates for mildly ill-posed inverse problems.

4.2. Improved convergence rates for Poisson and empirical process data

In section 4.1 we recalled the known convergence rate results for variational regularization methods of nonlinear ill-posed inverse problems with Poisson and empirical process data, which have been intensively studied in many other papers [6, 14, 16, 18, 37]. However, the convergence results of Theorem 4.11 in [18] for Poisson data or Theorem 4.6 in [6] for data given by empirical processes are shown to be not optimal in expectation. In this subsection we will further study a convergence analysis of the estimator (2) and improve the reconstruction error bounds in expectation in terms of some noise level for the variational regularization method with Poisson and empirical process data. To derive a better convergence rate for these data, we first need to bound the noise error **err** and we have to show some technical lemmas of composition operators.

Lemma 4.5. *Let* $\mathbb{M} \subset \mathbb{R}^d$ *be a bounded Lipschitz domain. Suppose that* $g \in H_2^s(\mathbb{M})$ *with* $s \geq \frac{d}{2}$ *and* $g \geq a > 0$, *then* $\ln(g) \in H_2^s(\mathbb{M})$ *and* $g^{-1} \in H_2^s(\mathbb{M})$. *Moreover, there exists a constant* $\tilde{C} > 0$ *such that*

$$\|\ln(g)\|_{H_2^s(\mathbb{M})} \leq \tilde{C} \|g - 1\|_{H_2^s(\mathbb{M})} \tag{36}$$

holds trues for $\|g - 1\|_{H_2^s(\mathbb{M})} \leq R$ *with* $R > 0$. *Moreover,*

$$\left\| \frac{1}{g} \right\|_{H_2^s(\mathbb{M})} \leq \tilde{C} \left(\frac{1}{a} \|\mathbf{1}\|_{H_2^s(\mathbb{M})} + \|g - a\|_{H_2^s(\mathbb{M})} \right) \tag{37}$$

for all $g \geq a$ *with* $\|g\|_{H_2^s(\mathbb{M})} \leq R$, *where* $\mathbf{1}$ *denotes the constant function with value* 1.

Proof. Set $v := g - 1$ and define a composition operator $T_H(v) = H \circ v$ on Sobolev spaces $H_2^s(\mathbb{M})$ associated with a function $H : \mathbb{R} \to \mathbb{R}$ given by $H(x) = \ln(x + 1)$ for $x \in [a + 1, \infty)$. Then (36) is equivalent to

$$\|H \circ v\|_{H_2^s(\mathbb{R}^d)} \leq \widetilde{C} \|v\|_{H_2^s(\mathbb{R}^d)} \tag{38}$$

for all $v \in H_2^s(\mathbb{M})$ satisfying $v + 1 \in [a, \infty)$ and $\|v\|_{H_2^s(\mathbb{R}^d)} < R$. The function H can be extended from the interval $[a - 1, \infty)$ to a function in $C^{m+1}(\mathbb{R})$ and $H(0) = 0$. Hence, by applying Theorem A in Adams and Frazier [1] to the cases $H \circ v$, we have

$$\|H \circ v\|_{H_2^s(\mathbb{R}^d)} \leq c \|H\|_{C^{m+1}} \left(\|v\|_{H_2^s(\mathbb{R}^d)} + \|v\|_{H_2^s(\mathbb{R}^d)}^{\tilde{s}} \right) \tag{39}$$

for all $v \in H_2^s(\mathbb{R}^d)$ with $\tilde{s} = \max(1, s)$ and for $c > 0$. Then the inequality (39) shows (38) with constant $\tilde{C} =: c \|H\|_{C^{m+1}} (1 + R^{\tilde{s}-1})$. In the case $\mathbb{M} \neq \mathbb{R}^d$, we use that $H_2^s(\mathbb{M}) = \{f|_{\mathbb{M}} : f \in H_2^s(\mathbb{R}^d)\}$ with norm by the infimum over $\|f\|_{H_2^s(\mathbb{R}^d)}$. As $(H \circ v)|_{\mathbb{M}} = H \circ v|_{\mathbb{M}}$, then we obtain the assertion for general $\mathbb{M}$.

Similarly, to show the second argument let $H(x) = \frac{1}{x+a} - \frac{1}{a}$ for $a > 0$ and then the function $H : \mathbb{R} \to \mathbb{R}$ such that $H(0) = 0$. Hence, by applying Theorem A in Adams and Frazier [1] once again, we derive (37). $\square$

Assumption 4.6 (Smoothing and Lipschitz properties). *Let* $\mathbb{M} \subset \mathbb{R}^d$ *be a bounded Lipschitz domain. Let* $F : \mathcal{B} \subset \mathbb{X} \to B_{2,2}^u$ *be Lipschitz continuous for some* $u > d/2 > 0$, *i.e.,*

$$\frac{1}{L} \|f_1 - f_2\|_{\mathbb{X}} \leq \|F(f_1) - F(f_2)\|_{B_{2,2}^u(\mathbb{M})} \leq L \|f_1 - f_2\|_{\mathbb{X}}$$

for all $f_1, f_2 \in \mathcal{B}$ *and for some Lipschitz constant* $L > 0$.

It is necessary to emphasize that the smoothness assumption implies the boundedness of F from $\mathbb{X}$ to Besov space $B_{2,2}^s(\mathbb{M})$ if $F : \mathbb{X} \to B_{2,2}^s(\mathbb{M})$ is linear and $\mathbb{X}$ is a Hilbert space and the penalty term is quadratic.

Proposition 4.7. *Suppose that Assumption 4.6 holds true and* $F(f) \geq 0$ *for all* $f \in \mathcal{B} \subset \mathbb{X}$ *and* $F(\mathcal{B})$ *is bounded in* $B_{2,2}^s(\mathbb{M})$ *such that* $\sup_{f \in \mathcal{B}} \|F(f)\|_{B_{2,2}^s(\mathbb{M})} < \infty$ *with some* $s \in [d/2, u)$. *Moreover, let* $\widehat{f}_\alpha$ *be a minimizer of Tikhonov functional (2), then there exists a constant* $\widetilde{C}_{con} > 0$ *such that*

$$\left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,1}^s(\mathbb{M})} \leq \widetilde{C}_{con} \|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^{\frac{s}{2u}} \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}} \tag{40}$$

for all $g = F(\widehat{f}_\alpha) \in B_{2,1}^s(\mathbb{M})$.

Proof. To bound the logarithmic function in (42), we will first use the interpolation inequality in Besov spaces with the domain $\mathbb{M}$. By Theorem 1.33 in Triebel [31], we have

$$\left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B^s_{2,1}(\mathbb{M})} \leq C_1 \left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|^{\frac{u}{s}}_{B^u_{2,2}(\mathbb{M})} \left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|^{1 - \frac{u}{s}}_{B^0_{2,2}(\mathbb{M})} \tag{41}$$

for some constant $C_1 > 0$ and $u > s > 0$. Note that $B^u_{2,2}(\mathbb{M}) = H^u_2(\mathbb{M})$ and $B^0_{2,2}(\mathbb{M}) = L^2(\mathbb{M})$ for the smoothness $u > s > 0$. As a consequence of Lemma 4.5 and by Theorem 2.8.3 in Triebel [30] for the bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$, the first factor on the right hand side of (41) can be bounded by

$$\left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B^u_{2,2}(\mathbb{M})} \leq C_2 \left(\frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \right)^m \left\| \frac{F(\widehat{f}_\alpha) - g^\dagger}{g^\dagger + \delta} \right\|_{B^u_{2,2}(\mathbb{M})}$$

$$\leq C_2 \left(\frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \right)^m \left\| \frac{1}{g^\dagger + \delta} \right\|_{B^u_{2,2}(\mathbb{M})} \left\| F(\widehat{f}_\alpha) - g^\dagger \right\|_{B^u_{2,2}(\mathbb{M})} \tag{42}$$

with some constant $C_2 > 0$, where we used the inequality $\frac{g + \delta}{g^\dagger + \delta} \geq \frac{\delta}{2(\|g^\dagger\|_{L^\infty} + \delta)}$ for small $\delta \leq 1$. The second factor on the right hand side of (42) be bounded based on the result (37) from Lemma 4.5. Moreover, by Assumption 4.6, we have

$$\|F(\widehat{f}_\alpha) - g^\dagger\|_{B^u_{2,2}(\mathbb{M})} \leq L \|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}} \tag{43}$$

for Lipschitz constant $L > 0$. From the inequality (42) and (43) we have

$$\left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B^u_{2,2}(\mathbb{M})} \leq C_{\delta, g^\dagger, L} \left\| \widehat{f}_\alpha) - f^\dagger \right\|_{\mathbb{X}} \tag{44}$$

holds true for some constant $C_{\delta, g^\dagger, L} > 0$, which depends on $g^\dagger, \delta > 0$ and Lipschitz constant $L > 0$.

Now we will further bound the second factor on the right hand side of (40). Similarly, with the help of Lemma 4.5 we have

$$\left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B^0_{2,2}(\mathbb{M})} \leq \frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \left\| \frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} - 1 \right\|_{L^2(\mathbb{M})}$$

$$\leq \frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \left\| \frac{1}{g^\dagger + \delta} \right\|_{L^\infty(\mathbb{M})} \left\| F(\widehat{f}_\alpha) - g^\dagger \right\|_{L^2(\mathbb{M})}$$

$$= C_{\delta, g^\dagger} \left\| F(\widehat{f}_\alpha) - g^\dagger \right\|_{L^2(\mathbb{M})} \tag{45}$$

for some constant $C_{\delta, g^\dagger} = \frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \left\| \frac{1}{g^\dagger + \delta} \right\|_{L^\infty(\mathbb{M})} > 0$ depending on δ and $g^\dagger$, where $\|g^\dagger\|_{L^\infty}$ can be seen as a constant function. Then by applying Lemma 2.6 [18] for the norm $\|F(\widehat{f}_\alpha) - g^\dagger\|_{L^2(\mathbb{M})}$, we obtain

$$\|F(\widehat{f}_\alpha) - g^\dagger\|_{L^2(\mathbb{M})} \leq \left(\frac{3}{2} \|g^\dagger\|_{L^\infty(\mathbb{M})} + \frac{4}{3} \|F(\widehat{f}_\alpha)\|_{L^\infty(\mathbb{M})} + 2\delta \right)^{\frac{1}{2}} \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}, \tag{46}$$

where $\|F(\widehat{f}_\alpha)\|_{L^\infty(\mathbb{M})}$ is bounded due to the fact that $\sup_{f \in \mathscr{B}} \|F(\widehat{f}_\alpha)\|_{B^s_{2,2}(\mathbb{M})}$ is bounded with some $s \in [\frac{d}{2}, u]$. Thus the first factor can be seen as a constant function which we denote by

$$C_3 = \left(\frac{3}{2} \|g^\dagger\|_{L^\infty(\mathbb{M})} + \frac{4}{3} \|F(\widehat{f}_\alpha)\|_{L^\infty(\mathbb{M})} + 2\delta \right)^{\frac{1}{2}} > 0.$$

Therefore (46) is further bounded by

$$\|F(\widehat{f}_\alpha) - g^\dagger\|_{L^2(\mathbb{M})} \leq C_3\, \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}. \tag{47}$$

By plugging (47) into (45) we have

$$\left\|\ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta}\right)\right\|_{B^0_{2,2}(\mathbb{M})} \leq C_{\delta,g^\dagger,C_3}\, \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}. \tag{48}$$

Finally, by inserting (44) and (48) into (41) we obtain

$$\left\|\ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta}\right)\right\|_{B^s_{2,1}(\mathbb{M})} \leq C_1 C_2 C_{\delta,g^\dagger,L} C_{\delta,g^\dagger,C_3} \|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^{\frac{s}{u}} \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}}.$$

Setting $\widetilde{C}_{con} := C_1 C_2 C_{\delta,g^\dagger,L} C_{\delta,g^\dagger,C_3} > 0$, which yields the assertion. $\qquad\square$

Theorem 4.8. *Let the assumptions of Proposition 4.7 are satisfied and Assumption 4.1 holds true. Let $\widehat{f}_\alpha$ be a minimizer of (2). Then there exists some constant $\widetilde{C}_{con} > 0$ such that*

$$\mathbf{err}(F(\widehat{f}_\alpha)) \leq 2^{\frac{s}{u}}\widetilde{C}_{con}^2 \alpha^{-\frac{s}{u}} \left\|G_n - g^\dagger\right\|_{B^{-s}_{2,\infty}}^2 + \alpha\Psi(2\alpha) \tag{49}$$

and

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq 2^{\frac{s}{u}}\widetilde{C}_{con}^2 \alpha^{-1-\frac{s}{u}} \left\|G_n - g^\dagger\right\|_{B^{-s}_{2,\infty}}^2 + \Psi(2\alpha) \tag{50}$$

for all $\alpha > 0$ and for $s \geq \frac{d}{2}$.

Proof. By using Hölder's inequality and Theorem 2.6.1 in Triebel [29] and Proposition 4.7, we bound the error term (29) by

$$\mathbf{err}(F(\widehat{f}_\alpha)) = \left|\int_{\mathbb{M}} \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta}\right) (\mathrm{d}G_n - g^\dagger\mathrm{d}x)\right|$$

$$\leq \|G_n - g^\dagger\|_{B^{-s}_{2,\infty}(\mathbb{M})} \left\|\ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta}\right)\right\|_{B^s_{2,1}(\mathbb{M})}$$

$$\leq \widetilde{C}_{con}\|G_n - g^\dagger\|_{B^{-s}_{2,\infty}(\mathbb{M})}\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^{\frac{s}{u}} \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}}, \tag{51}$$

where we define

$$\|G_n - g^\dagger\|_{B^{-s}_{2,\infty}(\mathbb{M})} := \sup_{\|\mathfrak{f}\|_{B^s_{2,1}(\mathbb{M})} \leq 1} \left|\int_{\mathbb{M}} \mathfrak{f}(x)(\mathrm{d}G_n - g^\dagger\mathrm{d}x)\right|$$

and $B^{-s}_{2,\infty}(\mathbb{M})$ is a dual space of $B^s_{2,1}(\mathbb{M})$. By monotonicity of Ψ and (31), we have

$$\frac{1}{2}\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^2 \leq \frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + \Psi(\alpha) \leq \frac{1}{\alpha}\left(\mathbf{err}(F(\widehat{f}_\alpha)) + \alpha\Psi(2\alpha)\right). \tag{52}$$

By plugging the deterministic error bound (30) in Theorem 4.2 and (52) into (51) and using Young's inequality $xy \leq \frac{1}{p}x^p + \frac{1}{q}y^q$ for all $x, y > 0$ and $p, q > 0$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$, the noise error (51) is again bounded by

$$\mathbf{err}(F(\widehat{f}_\alpha)) \leq \widetilde{C}_{con}\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^{\frac{s}{u}} \mathscr{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}}\|G_n - g^\dagger\|_{B^{-s}_{2,\infty}}$$

$$\leq \widetilde{C}_{con} 2^{\frac{s}{2u}}(2\alpha)^{\frac{u-s}{2u}}\left(\frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + \Psi(2\alpha)\right)^{\frac{1}{2}}\|G_n - g^\dagger\|_{B^{-s}_{2,\infty}}$$

$$\leq \widetilde{C}_{con} 2^{\frac{s}{2u}} (2\alpha)^{\frac{u-s}{2u}} \left(\frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} \right)^{\frac{1}{2}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}$$

$$+ \widetilde{C}_{con} 2^{\frac{s}{2u}} \alpha^{\frac{u-s}{2u}} \Psi(2\alpha)^{\frac{1}{2}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}$$

$$\leq \frac{1}{2}\mathbf{err}(F(\widehat{f}_\alpha)) + \frac{\alpha}{2}\Psi(2\alpha) + \frac{1}{2} 2^{\frac{s}{u}} \alpha^{\frac{u-s}{u}} \left(\frac{\widetilde{C}_{con}}{\sqrt{\alpha}} \right)^2 \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2$$

$$+ \frac{1}{2} 2^{\frac{s}{u}} \alpha^{\frac{u-s}{u}} \left(\frac{\widetilde{C}_{con}}{\sqrt{\alpha}} \right)^2 \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2, \tag{53}$$

by subtracting $\frac{1}{2}\mathbf{err}(F(\widehat{f}_\alpha))$ on the both side of the inequality (53), we obtain

$$\mathbf{err}(F(\widehat{f}_\alpha)) \leq 2^{\frac{s}{u}} \widetilde{C}_{con}^2 \alpha^{-\frac{s}{u}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2 + \alpha\Psi(2\alpha).$$

Finally, by inserting this inequality into (49), we derive (50). $\qquad \square$

Remark 4.9. *(i) Similarly, we can also obtain the optimal bound for the discrete type of Poisson and empirical noise as in Theorem 4.8. In discrete setting, the observed quatities will be discrete measurement $\mathbf{g}^{obs} \in \mathbb{N}^J$ in practice. If the measurement manifold $\mathbb{M} = \bigcup_{j=1}^{J} \mathbb{M}_j$ with $|\mathbb{M}_j| > 0$ for all $j \in J$, then the discrete measurements are given by $\mathbf{g}^{obs} = G_n(\mathbb{M}_j)$ and $\mathbf{g}^\dagger = \int_{\mathbb{M}_j} g^\dagger dx$. Thus, the effective noise level is given by*

$$\mathbf{err}_J(\mathbf{g}) := \left| \sum_{j=1}^{J} \ln\left(\frac{\mathbf{g} + \delta}{\mathbf{g}^\dagger + \delta} \right) (\mathbf{g}^{obs} - \mathbf{g}^\dagger) \right|$$

for all $\mathbf{g} \in \mathbb{N}^J$ and shift parameter $\delta > 0$, see the more detail about the discretized case in [16].

(ii) Obviously, it is possible to extend the the error bounds (49) and (50) into Banach space setting. However, we only consider here the quadratic penalty functional, i.e., $\mathscr{R}(f) = \frac{1}{2}\|f\|_{\mathbb{X}}^2$. The same results also hold true for Banach spaces with the norm powers $\mathscr{R}(f) = \frac{1}{q}\|f\|_{\mathbb{X}}^q$ for $q > 2$.

Theorem 4.10. *Let $F : \mathscr{B} \subset L^2(\mathbb{M}) \to B_{2,1}^u(\mathbb{M})$ be a continuous operator and $F(f) \geq 0$ for all $f \in \mathscr{B}$. Assume that $F(\mathscr{B})$ is bounded in $B_{2,1}^s(\mathbb{M})$ and $\sup_{f \in \mathscr{B}} \|F(f)\|_{B_{2,1}^s(\mathbb{M})} < \infty$ with some $s \in [d/2, u)$. Suppose that Assumptions 4.1 and 4.6 hold true with the norm topology for index $u > \frac{d}{2} > 0$ and let $\widehat{f}_\alpha$ be a global minimizer. Moreover, set $\tau \in (0, u)$ and $\tilde{d} \geq d$. Then there exist some constant $C_1, C_2, C_3 > 0$ such that for all $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ with $\|f^\dagger\|_{B_{2,\tau}^\tau(\mathbb{M})} \leq \rho$ and for the parameter choice rule*

$$\alpha^* = (\rho\sqrt{n})^{-\frac{2\tau}{u + \tilde{d}/2 + \tau}}$$

the error bound

$$\left\| \widehat{f}_\alpha - f^\dagger \right\|_{\mathbb{X}}^2 \leq C_1 \rho^{\frac{4\tau + 2\tilde{d}}{2\tau + 2a + \tilde{d}}} n^{-\frac{2\tau}{2\tau + 2u + \tilde{d}}} \left(1 + n\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2 \right) \tag{54}$$

holds for all $n \in \mathbb{N}$. Then we obtain

$$\mathbf{E}\left[\left\| \widehat{f}_{\alpha^*} - f^\dagger \right\|_{\mathbb{X}}^2 \right]^{1/2} \leq C_2 \rho^{\frac{2\tau + \tilde{d}}{2\tau + 2u + \tilde{d}}} n^{-\frac{\tau}{2\tau + 2u + \tilde{d}}} \tag{55}$$

as $n \to \infty$. Moreover, if Conjecture 3.2 holds true with $\tilde{d} = d$, then we obtain the following convergence rate

$$\mathbf{E}\left[\left\| \widehat{f}_{\alpha^*} - f^\dagger \right\|_{\mathbb{X}}^2 \right]^{1/2} \leq C_3 (\ln n)^m \rho^{\frac{2\tau + 2d}{2\tau + 2u + d}} n^{-\frac{\tau}{2\tau + 2u + d}} \tag{56}$$

as $n \to \infty$, which is order optimal if $m = 0$.

Proof. By combining the results of Lemma 4.3 and Theorem 4.8, we have

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq 2^{\frac{s}{u}}\widetilde{C}_{con}^2\alpha^{-1-\frac{s}{u}}\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})} + c\rho^2\alpha^{\frac{\tau}{u}}.$$

The parameter choice $\alpha \sim \rho^{-\frac{2u}{2u+2\tau+\tilde{d}/2}}n^{-\frac{u}{u+\tilde{d}/2+\tau}}$ with $s = \tilde{d}/2 > 0$ gives the following error bound

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq C_1\rho^{\frac{4\tau+2\tilde{d}}{2\tau+2a+d}}n^{-\frac{2\tau}{2\tau+2u+\tilde{d}}}\left(1 + n\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2\right) \tag{57}$$

for all $n \in \mathbb{N}$ and some constant $C_1 > 0$. Finally, by taking expectation on (57) and Theorem 3.3 we find that

$$\sup_{n\in\mathbb{N}}\mathbf{E}\left[n\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2\right] < \infty$$

if $\tilde{d} > d$ or if Conjecture 3.2 holds true with $m = 0$ which implies (54) and (55) for $m = 0$. If $m > 0$, then

$$\sup_{n\in\mathbb{N}}\mathbf{E}\left[\frac{n}{(\ln n)^m}\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2\right] < \infty$$

and we derive (56) for $m > 0$.

Similarly, by combining the deviation inequality (17) of Theorem 3.3 for empirical process and Theorem 4.8 with (55) we can also derive the same convergence rate as shown in (55), which complete the proof of the theorem. $\square$

On the one hand, the results of Theorem 4.10 show that convergence rates for variational regularization methods are close to optimal order compare to rates for wavelet Galerkin approximation methods studied by Antoniadis and Bigot [2]. If our conjecture (3.2) holds true, then Theorem 4.10 would establish with the exponent $\frac{\tau}{2\tau+2u+d}$, which is known to be order optimal in minimax setting. In the variational method, the convergence result is not affected by the nonlinearity of the forward operator, however Antoniadis and Bigot in [2] are restricted to linear operators in certain function spaces.

On the other hand, the convergence rate result (55) is more explicit and close to optimal rate than the convergence result derived by Werner and Hohage [37], their convergence rate is strictly smaller than the optimal order $\frac{\tau}{2\tau+2u+d}$. Furthermore, here we study the convergence analysis for variational regularization of inverse problems with Poisson and empirical process data simultaneously.

Acknowledgements

Reference

[1] D. R. Adams and M. Frazier. Composition operators on potential spaces. *Proceedings of the American Mathematical Society*, 114 (1): 155-165, 1992.

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

[3] M. Benning and M. Burger. Error estimates for general fidelities. *Electronic Tansactions on Numerical Analysis*, 38:44–68, 2011.

[4] M. Bertero, P. Boccacci, G. Desiderà, and G. Vicidomini. Image deblurring with poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006, 2009.

[5] L. Cavalier and J.-Y. Koo. Poisson intensity estimation for tomographic data using a wavelet shrinkage approach. *IEEE Transactions on Information Theory*, 48 (10):2794–2802, 2002.

[6] F. Dunker and T. Hohage. On parameter identification in stochastic differential equations by penalized maximum likelihood. *Inverse Problems*, 30 (9):095001, 2014.

[7] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. Society for Industrial and Applied MAthematics, 1999.

[8] J. Flemming. *Generalized Tikhonov regularization and modern convergence rate theory in Banach space*, Shaker Verlag, Aachen, 2012.

[9] J. Flemming, B. Hofmann Convergence rates in constrained Tikhonov regularization: equivalence of projected source conditions and variational inequalities. *Inverse Problems*, 27 (8): 085001, 2011.

[10] J. Flemming, B. Hofmann and P. Math. Sharp converse results for the regularization error using distance functions. *Inverse Problems*, 27 (2): 025006, 2011.

[11] L. Grafakos. *Classical Fourier Analysis*, Springer New York, 2009.

[12] M. Grasmair. Generalized Bregman distances and convergence rates for non-convex regularization methods, *Inverse Problems*, 26 (11): 115014, 2010.

[13] H. T. Hans-Jurgen Schmeisser. *Topics in Fourier Analysis and Function Spaces*, First Edition, Wiley, 1987.

[14] B. Hofmann, B. Kaltenbacher, C. Pschl and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23 (3), 987-1010, 2007.

[15] T. Hohage and F. Weidling Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM Journal of Numerical Analysis*, 55 (2): 598-620, 2017.

[16] T. Hohage and F. Werner. Iteratively regularized newton-type methods for general data misfit functionals and applications to poisson data. *Numerische Mathematik*, 123 (4):745–779, 2013.

[17] T. Hohage and F. Werner. Convergence rates for inverse problems with impulsive noise. *SIAM Journal on Numerical Analysis*, 52 (3):1203–1221, 2014.

[18] T. Hohage and F. Werner. Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, 32 (9):093001, 2016.

[19] T. Hohage and Philip Miller. Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces. *Inverse Problems*, 35 (6): 065005, 2019.

[20] J. F. C. Kingman. *Poisson processes*, Volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1993.

[21] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *The Annals of Probability*, 28 (2): 863–884, 2000.

[22] M. Moussai. On the composition of functions in multidimensional Besov spaces. *Mathematical Inequalities & Applications*, 2: 501–514, 2017.

[23] A. Munk and M. Pricop. On the self-regularization property of the EM algorithm for poisson inverse problems. 431–448, 2009.

[24] R. Nowak and E. Kolaczyk. A statistical multiscale framework for poisson inverse problems. *IEEE Transactions on Information Theory*, 46(5):1811–1825, 2000.

[25] P. Reynaud-Bouret. Adaptive estimation of the intensity of inhomogeneous poisson processes via concentration inequalities. *Probability Theory and Related Fields*, 126(1):103–153, 2003.

[26] B. Sprung. Upper and lower bounds for the Bregman divergence. *Journal of Inequalities and Applications*, 4(1), 2019.

[27] Z. Szkutnik. Unfolding intensity function of a poisson process in models with approximately specified folding operator. *Metrika*, 52(1):1–26, 2000.

[28] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*. 126 (3): 505-563, 1996.

[29] H. Triebel *Interpolation theory, Function spaces, Differential operators*. North-Holland, Amsterdam, 1978.

[30] H. Triebel. *Theory of Function Spaces*. Springer Basel, 1983.

[31] H. Triebel. Function Spaces and Wavelets on Domains. EMS Tracts in Mathematics 7. European Mathematical Society, 2008.

[32] A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer New York, 2009.

[33] Y. Vardi, L.A. Shepp and L. Kuafman. A Statistical Model for Positron Emission Tomography. *Journal of the American Statistical Association*, 80 (389): 8-20, 1985.

[34] M. C. Veraar. Regularity of Gaussian White noise on the *d*-dimensional torus. *Banach Center Publications*, 95: 385-398, 2011.

[35] F. Weidling and T. Hohage. Variational source conditions and stability estimates for inverse electromagnetic medium scattering problems. *Inverse Problems and Imaging*, 11(1), 2017.

[36] F. Weidling, B. Sprung and T. Hohage. Optimal convergence rates for Tikhonov regularization in Besov spaces. *Preprint. arxiv.org/abs/1803.11019*, 2018.

[37] F. Werner and T. Hohage. Convergence rates in expectation for Tikhonov-type regularization of inverse problems with poisson data. *Inverse Problems*, 28(10):104004, 2012.

[38] Z. B Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications*, (157) 189–210, 1991.