Convergence rates for variational regularization of inverse problems in exponential families

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen

Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität zu Göttingen

im Promotionsprogramm

"PhD School of Mathematical Sciences (SMS)"

der Georg-August University School of Science (GAUSS)

vorgelegt von

Yusufu Simayi

aus Xinjiang, China

Göttingen, 2019

Betreuungsausschuss:

Prof. Dr. Thorsten Hohage

Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Prof. Dr. Axel Munk

Institut für Mathematische Stochastik, Universität Göttingen

Mitglieder der Prüfungskommission:

Referent:

Prof. Dr. Thorsten Hohage

Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Korreferent:

Prof. Dr. Axel Munk

Institut für Mathematische Stochastik, Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Gerlind Plonka-Hoch

Institut für Numerische und Angewandte Mathematik, Universität Göttingen

Prof. Dr. Russell Luke

Institut für Mathematische Stochastik, Universität Göttingen

Dr. Frank Werner

Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Dominic Schuhmacher

Institut für Mathematische Stochastik, Universität Göttingen

Tag der mündlichen Prüfung: 12.9.2019.

Declaration of Authorship

I, Yusufu SIMAYI (Yusup Ismayil), declare that this thesis titled, "Convergence rates for variational regularization of inverse problems in exponential families" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at University of Göttingen.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at University of Göttingen or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others and I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: 12.09. 2019

"Mathematics is a game played according to certain simple rules with meaningless marks on paper."

David Hilbert

Acknowledgements

It is my great pleasure to express my gratitude to all of my colleagues and friends. First of all, I would like to thank Prof. Thorsten Hohage for the supervision of my thesis. He strongly supported me to finish my thesis and gave several helpful hints and suggestions. He was not only guided me into the very interesting area of statistical inverse problems and regularization theory, but also suggested the broad subject for my thesis. Moreover, I would like to thank Prof. Axel Munk as co-referee for this thesis and gave some fruitful suggestion about my future research interest. To complete my work during these years, I really faced some difficult problems and it was not possible to achieve my work without some important hints from Prof. Thorsten Hohage. Especially, I got stuck several times to bound the expectation of Poisson and empirical process and large deviation inequalities for such processes in negative Besov norms.

Furthermore, I also would like to express my gratitude to the other members of the inverse problems working group in Göttingen, they pointed out several remarks about my work. I am really happy to stay in a good working environment at the University of Göttingen. Especially, I am very much indebted to my colleague Benjamin Sprung for the helpful idea on the proof of Theorem 5.14, and Frank Werner who is a former member of the inverse problems group, his research work provided me some basic tools to achieve my work on this research project. During my whole PhD studies I was supported by the Research Training Group 2088 **"Inverse problems in exponential families: likelihood data fidelities B1"** and I would like to acknowledge not only the financial support but also many helpful international conferences and the exchange of ideas and information with other members of the RTG 2088 during the annual workshops.

Finally, I really appreciate my wife Arezugul Mamat for her tireless support and encouraging help during the whole three years. She is taking care of my kid Babur after she left Germany in 2017. Although we didn't see each other after her return to China in 2017, she has always shown her patience and supported me to finish my work successfully in Göttingen. Honestly speaking, I really miss my parents and my families in China during my PhD study, unfortunately I was not able to go back to China due to a harsh political environment in my region.

Contents

List of Symbols

$\mathbb{X}$	Hilbert space containing causes of an inverse problem
$\mathbb{Y}$	Hilbert space containing effects of an inverse problem
$\mathbb{T}^d$	d-dimensional torus
$\mathscr{D}(\mathbb{R}^d)$	space of all complex-valued, infinitely differentiable functions on $\mathbb{R}^d$
$\mathscr{D}'(\mathbb{R}^d)$	the topological dual space of $\mathscr{D}(\mathbb{R}^d)$
$\mathbb{M}$	Riemannian manifold
$H^s([0, 2\pi])(\mathbb{M})$	periodic Sobolev space in $[0, 2\pi]$ with smoothness index s
$W_p^s(\mathbb{M})$	Sobolev spaces with parameters s, p
$B_{p,q}^s(\mathbb{M})$	Besov spaces with parameters s, p, q
F	nonlinear operator of an inverse problem, $F : \mathbb{X} \to \mathbb{Y}$
T	linear operator of an inverse problem, $T : \mathbb{X} \to \mathbb{Y}$
T^*	adjoint operator of T, $T^* : \mathbb{Y}^* \to \mathbb{X}^*$
α	regularization parameter
α_k	regularization parameter at a kth iteration
q_α	filter function depending on some regularization parameter α
$\tau_{\mathbb{X}}, \tau_{\mathbb{Y}}$	topologies on the spaces $\mathbb{X}$ and $\mathbb{Y}$
$(\Omega, \mathcal{F}, \mathbf{P})$	probability space with domian Ω, sigma field $\mathcal{F}$ and probability $\mathbf{P}$
$f^\dagger$	true solution of an inverse problem
$g^\dagger$	exact data of an inverse problem, $g^\dagger = F(f^\dagger)$
g^{obs}	observed data of an inverse problem
G_n	temporally normalized Poisson and empirical process
n	exposure time
δ	a deterministic noise level
ε	a stochastic noise level
W	white noise process
δ_x	Dirac measures at point x
$\mathcal{S}$	data fidelity functional
$\mathcal{T}$	exact data fidelity functional
$\mathcal{R}$	penalty functional
$\mathbf{T}(f)$	Tikhonov functional with element f

$\partial \mathcal{R}(f)$	subdifferential of $\mathcal{R}$ at $f \in \mathbb{X}$
μ	drift coefficient
σ	volatility or diffusion
$\widehat{f}_\alpha$	penalized maximum likelihood estimator in $\mathbb{X}$
$\widehat{f}_k$	kth iteration of a penalized maximum likelihood estimator in $\mathbb{X}$
$\mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger)$	Bregman divergence of f and $f^\dagger$ w.r.t. $f^* \in \partial \mathcal{R}(f)$
$KL(g, g^\dagger)$	Kullback-Leibler divergence between g and $g^\dagger$
$\mathbf{err}, \mathbf{err}_k$	effective noise levels
Φ, Ψ	index functions
Φ^*	Fenchel conjuagtion of index function Φ
ρ_Θ	Orlicz functions

Chapter 1

Introduction

Inverse problems are an increasingly important branch of mathematical science, which is relevant to many other scientific areas. These problems consist in determining some model parameters from observed data, while the direct or forward problems consist in predicting the data with given models. The latter problems are in general well-understood and they have stable solutions. From the mathematical point of view, the term *well-posed* was introduced by Hadamard [37]: A problem is called *well-posed*, if

  (a) there exists a solution to the problem,

  (b) there is at most one solution to the problem,

  (c) the solution depends continuously on the data.

If problems are not well-posed in the sense of Hadamard, the problems are called *ill-posed*. If one of two problems which are inverse to each other is ill-posed and the other is well-posed, we call it the *inverse problem* and the other one the *direct* or *forward problem*. Especially the condition (c) is the most delicate among these conditions. We always assume that the inverse problem can be formulated as an operator equation

$$F(f) = g, \tag{1.1}$$

where g are the data, f the unknown parameters and the forward (possibly nonlinear) operator $F : \mathbb{X} \to \mathbb{Y}$ describes the dependence between Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$. If the solution does not depend continuously on the data and the measured data is perturbed by some noise, then naive reconstruction are dominated by propagated data noise and must be considered useless. For the operator equations (1.1), the Hadamard's criteria (a)-(c) of *well-posedness* can be written as:

  (a) the operator is surjective,

  (b) the operator is injective, i.e., F is one to one,

  (c) the operator is continuously invertible, i.e., F^{-1} is continuous.

In this research project, we focus on the problem of convergence rates for variational regularization methods of linear and nonlinear mildly ill-posed inverse problems with stochastic data including additive random noise, Poisson data and data given by empirical processes, respectively. First of all, we analyze convergence results for linear inverse problems with additive random noise based on the spectral theory. Secondly, we further discuss inverse problems with other stochastic noise models described by exponential families. For instance, inverse problems with Poisson data arise in many photonic imaging modalities, engineering and astronomy (see [9, 51, 84]). The observed data will be modeled by a Poisson process with density

$ng^\dagger$ and n corresponds to the expected total number of points, which have interacted with an unknown quantity of interest $f^\dagger$. We denote the Poisson process by $\widetilde{G}_n$. The point counting data are Poisson distributed due to some physical and mathematical reasons as we describe in Chapter 2. In other words, we can observe a random set of points that corresponds to the particle detections and reconstruct the underlying reason for the true particle density $g^\dagger$. The parameter $n > 0$ is often interpreted as an exposure time of the measurement procedure. As for the parameter $n > 0$, Werner [94] explained in detail for some applications in photonic imaging including Positron Emission Tomography, coherent X-ray imaging, astronomical imaging and fluorescence microscopy in his PhD thesis, where the exposure time n is limited due to some special reasons. For example, one may expect the normalized data $G_n = \frac{1}{n}\widetilde{G}_n$ tend to the exact data $g^\dagger$ as $n \to \infty$. The parameter n has various physical meanings in a specific application, for more details we refer to the PhD thesis of Werner [94].

Additionally, we will also consider inverse problems with data given by empirical processes, where the observed data are described by independent and identically distributed random variables $X_j, j = 1, \cdots, n$. The observed data will be drawn from an empirical process with probability density $g^\dagger$ and we denote the observed data by G_n, i.e.,

$$G_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{X_j}, \quad j = 1, \cdots, n,$$

where δ_{X_j} are Dirac measures. More details will be presented in Section 2.2. Inverse problems with empirical process data described by stochastic differential equations appear in financial econometrics, physics, parameter identification problems and social sciences etc, (see [12, 77]). In particular, the parametric and nonparametric estimation in parameter identification problems for the drift and diffusion in ergodic models are widely studied.

The maximum likelihood method to solve the equation (1.1) consists in seeking an approximation $\widehat{f}_{\text{app}}$ of true solution $f^\dagger$ as

$$\widehat{f}_{\text{app}} \in \underset{f}{\arg\max}\,\mathbf{P}\left(G_n; F(f)\right), \tag{1.2}$$

where $F(f)$ is true density. The approach (1.2) is extensively applied by many authors to obtain estimators for distribution characteristics. The approximation $\widehat{f}_{\text{app}}$ is not changed after replacing the probability distribution $\mathbf{P}$ in (1.2) by its negative logarithm and substituting argmax by argmin, which we denote as **data fidelity functional**:

$$\mathcal{S}(G_n, F(f)) := -\ln\mathbf{P}(G_n; F(f)),$$

where $\mathcal{S}$ is convex functional for the cases of Poisson and empirical process data in its first argument and minimize the problems with simple way. Nevertheless this approach cannot provide us stable approximations of $f^\dagger$ in the sense that $\widehat{f}_{\text{app}}$ does not depend continuously on the data due to discontinuity of the inverse operator F. Therefore, we now recall well-known regularization methods in inverse problems such as Tikhonov regularization and iteratively regularized Newton-type methods for ill-posed inverse problems. To apply such regularization methods, we usually need an additional term to stablize the reconstruction procedure. More precisely, to construct a stable approximation to $f^\dagger$, one needs to stablize the functional by

adding a **penalty term** $\mathcal{R} : \mathbb{X} \to (-\infty, \infty]$ and then the minimizer $\widehat{f}_\alpha$ is given by

$$\widehat{f}_\alpha \in \underset{f \in \mathbb{X}}{\arg\min} \left[\mathcal{S}(G_n, F(f)) + \alpha \mathcal{R}(f) \right]. \tag{1.3}$$

Here $\mathcal{R}$ in general includes a priori knowledge about the unknown solution $f^\dagger$ and stablizes the reconstruction precedure, $\alpha > 0$ is a *regularization parameter* and it should be chosen in a proper way. Due to our special interest in this thesis, we restrict ourselves to quadratic penalty terms $\mathcal{R}$. The minimization problem (1.3) is called *Tikhonov-type regularization method* after the Russian mathematician Tikhonov [87]. He initially considered approximating solutions to (1.1) as minimizers of

$$f \mapsto \mathbf{T}(f), \quad \text{with } \mathbf{T}(f) = \frac{1}{2} \|G_n - F(f)\|_{\mathbb{Y}}^2 + \frac{\alpha}{2} \|f\|_{\mathbb{X}}^2, \tag{1.4}$$

the functional in (1.4) is called standard Tikhonov functional. The method (1.4) can only be available for data error with a Gaussian distribution and it is appropriate to use a quadratic Hilbert space norm $\| \cdot \|_{\mathbb{Y}}$ as data fidelity functional, this will be discussed in Section 3.2. The minimization problem (1.3) can be seen as a generalization of the Tikhonov regularization method and is not restricted quadratic norm functionals (see e.g. [32, 78, 75]).

One may ask some questions about the reconstruction procedure described by (1.3). For example, one wants to know about the existence of minimizer $\widehat{f}_\alpha$, whether the minimizer $\widehat{f}_\alpha$ depend continuously on the data or not and whether the minimizers $\widehat{f}_\alpha$ converge in expectation or in probability to the true solution as $n \to \infty$ for a proper choice of regularization parameter α depending on t or not etc. Naturely, more questions will be raised on optimal orders of convergence rate in expectation can be achieved by proper choice of parameter α depending on noise levels. Although the convergence rates in expectation for Tikhonov-type regularization of nonlinear inverse problems with Poisson and empirical process data have been studied by Werner and Hohage [53, 95] and Dunker and Hohage [24], respectively, they are not successful to achieve optimal rates of convergence. We want to derive the convergence results of the following form for Poisson and empirical process data

$$\mathbf{E}\left[\left\| \widehat{f}_\alpha - f^\dagger \right\|_{\mathbb{X}}^2 \right] = O\left(\Phi\left(\frac{1}{\sqrt{n}} \right) \right) \tag{1.5}$$

as $n \to \infty$, where $\Phi : [0, \infty) \to [0, \infty)$ is concave and monotonically increasing and $\Phi(0) = 0$. More details about these will be given in Chapter 3. Our aim in this research is to improve the reconstruction error bounds in expectation under some reasonable assumptions on $\mathcal{S}$ and $\mathcal{R}$ and ideally obtain optimal bounds in the minimax sense. To derive better convergence rates than the aforementioned authors' results, we try to sharpen the error bound with the help of large deviation inequalities for stochastic processes in negative Besov norms, which will be stated in Chapter 4. Such deviation inequalities play an important role to control the difference between $\mathcal{S}(G_n, F(f))$ and $\mathcal{S}(g^\dagger, F(f))$ in probability as $n \to \infty$.

For nonlinear operators F the minimization problem (1.3) is unfortunately nonconvex even if the data fidelity functional $\mathcal{S}(G_n, \cdot)$ and penalty function $\mathcal{R}$ are convex. Therefore, (1.3) may have many local minima and it is very difficult to calculate the global minimizer $\widehat{f}_\alpha$. In many practical applications such as inverse obstacle scattering problems without phase and semi-blind deconvolution problems the forward operator F is nonlinear and the functional in (1.3) is not convex in f. So, we may

seek a better regularization approach to handle it and overcome such difficulties with other types of regularization methods.

As we know, Newton's method is always one of the most useful approaches for solving nonlinear equations. For a nonlinear injective Fréchet differentiable mapping between its domain $\mathcal{D}(F) \subset \mathbb{X}$ into $\mathbb{Y}$, one can write the operator equation (1.1) by substituting F by its linear approximation in each Newton step, i.e.,

$$F(\widehat{f}_k) + F^{'}[\widehat{f}_k](f - \widehat{f}_k) = g,$$

where $F^{'}[\widehat{f}_k]$ denotes the Fréchet derivative of F at $\widehat{f}_k$. Then we are able to approximate true solution $f^{\dagger}$ iteratively as

$$\widehat{f}_{k+1} \in \arg\min \mathcal{S}\left(F(\widehat{f}_k) + F^{'}[\widehat{f}_k](f - \widehat{f}_k); G_n\right), \tag{1.6}$$

but we still have similar problems as we described for (1.2) and the iterates $\widehat{f}_{k+1}$ do not depend continuously on the data such that (1.6) is still ill-posed. To solve the nonlinear operator equation (1.1), we additionally need some kind of regularization approaches to stablize (1.6). This leads to iteratively regularized Newton-type method of the form

$$\widehat{f}_{k+1} \in \underset{f \in \mathbb{X}}{\arg\min} \left[\frac{1}{\alpha_k}\mathcal{S}(F(\widehat{f}_k) + F^{'}[\widehat{f}_k](f - \widehat{f}_k), G_n) + \alpha_k \mathcal{R}(f)\right], \tag{1.7}$$

where $\widehat{f}_0$ is some initial guess, $\mathcal{R}$ is again a penalty term incorporate a-priori knowledge and $\alpha_k > 0$ is a sequence of regularization parameters. The method (1.7) has substantial advantages compared to Tikhonov-type variational regularization (1.3). For the Newton-type method (1.7), we expect faster convergence behavior, however it is hard to determine the suitable stopping index. Hohage et al [24, 51] derived abstract convergence rates for the method (1.7) of nonlinear inverse problems with Poisson and empirical process data in terms of an index function Φ, i.e.,

$$\mathbf{E}\left[\left\|\widehat{f}_k - f^{\dagger}\right\|_{\mathbb{X}}^2\right] = O\left(\Phi\left(\frac{1}{\sqrt{n}}\right)\right) \tag{1.8}$$

as $n \to \infty$, where k is a stopping index. However, the convergence rate result (1.8) is not optimal as well. To improve the convergence rates for the Newton-type method (1.7), we will further study the convergence analysis under some proper conditions in Chapter 6 for Gaussian white noise and in Chapter 7 for Poisson and empirical process data, respectively.

There exists a large amount of literature on regularization methods for linear and nonlinear ill-posed problems with stochastic data. More specifically, inverse problems with Poisson data have been intensively studied by many authors [2, 19, 51, 53, 69, 85, 94, 95]. Among them, Antoniadis and Bigot [2] achieved optimal convergence rates for Poisson inverse problems by using Wavelet-Galerkin approximation methods, however, their results are restricted to linear operators. Cavalier [19] provides the best possible rate of convergence by using a wavelet-vaguelette decomposition. Furthermore, Hohage and Werner [51, 53, 95] in recent years discuss the convergence rates for Tikhonov- type regularization and iteratively regularized Newton-type methods of nonlinear inverse problems with Poisson data in Banach

space setting, unfortunately their results do not show optimal orders for mildly ill-posed problems. The Tikhonov-type regularization method (1.3) has been considered by Bardsley [6], Benning and Burger [8] and Flemming [31, 32] under deterministic noise assumptions.

Inverse problems with empirical process data described by Fokker-Planck equation have also been studied by Dunker and Hohage [24] who obtained general convergence results of the risk for generalized Tikhonov regularization and iteratively regularized Newton-type methods, but their results also cannot yield optimal convergence rates under Hölder-type source conditions. Furthermore, a statistical convergence analysis for nonlinear inverse problems with additive random noise described by general noise models has been explored by Bauer et al [7], Bissantz et al [11] and others, they achieved optimal rates for regularization methods of linear and nonlinear inverse problems based on spectral theory under general source conditions. Inverse problems for logarithmic source conditions in the quadratic Hilbert space cases has been considered by Hohage [45], where he proved optimal convergence rates for linear operators.

To the best of our knowledge, no optimal rates of convergence exist for variational regularization of nonlinear mildly ill-posed inverse problems in exponential family such as Gaussian, Poisson and empirical process data. To approach this goal, we expect to fill these gaps and further improve the reconstruction error bounds for Tikhonov-type regularization and iteratively regularized Newton-type method of mildly ill-posed inverse problems with the stochastic noise models described by exponential families.

The structure of this thesis is organized as follows: In **Chapter 2** we will recall several stochastic noise models where the measurement data described by distributions of stochastic random variables and random processes such as additive random noise, Poisson and empirical processes, respectively. Moreover, we will present concentration inequalities correspond to those processes achieved by previous authors. Those inequalities provide the main tools for our analysis under stochastic noise assumptions. Furthermore, some important applications of these inverse problems with stochastic data will be presented.

In **Chapter 3** we will study optimal rates of convergence for regularization methods of linear inverse problems based on the spectral theory and variational analysis with Gaussian random noise as well as introduce some source conditions and necessary assumptions. First of all, we will recall the convergence results for general spectral regularization under deterministic and stochastic noise assumptions achieved by Bissantz et al [11]. Secondly, we are also able to show optimal convergence rates for variational regularization under Hölder-type source conditions, these convergence rates that we obtained by the variational regularization same as the rates for the regularization method based on the general spectral theory. Finally, the substantial advantages of variational regularization theory over methods based on the spectral theory will be presented.

In **Chapter 4** we will study deviation inequalities for empirical and Poisson processes in negative Besov norms. First of all, we will give a precise definition for an important class of function spaces including Fourier-based and wavelet-based Besov spaces and some fundamental properties. Then, we will show large deviation inequalities for the suprema of Poisson and empirical processes in negative Besov norms based on uniform concentration inequalities derived by Massart [65] and Reynaud-Bouret [76], respectively. Such a deviation inequality for Gaussian white noise has been derived by Veraar [92]. To improve the reconstruction error in

expectation between the regularized estimator $\widehat{f}_\alpha$ and $f^\dagger$, the results of this chapter play an essential role to derive optimal rates of convergence in the stochastic setting.

In **Chapter 5** we will deal with a statistical convergence analysis of variational regularization method (1.3) for mildly ill-posed problems with Poisson and empirical process data. The general spectral regularization methods are not applicable for inverse problems with Poisson and empirical process data due to the nonquadratic data fidelity term $\mathcal{S}$. In order to derive improved convergence bound under some statistical noise assumptions, we will first present deterministic convergence results for variational regularization with Poisson and empirical process data. Then, with the help of deviation inequalities that we derived in Chapter 4, we are able to improve the convergence rates for (1.5) with an explicit exponent.

In **Chapter 6** tangential cone conditions for nonlinear operators and the known convergence results of iteratively regularized Gauss-Newton method (IRGNM) will be stated under both deterministic and stochastic setting, where the data fidelity term $\mathcal{S}$ and penalty term $\mathcal{R}$ are given by squared norm in Hilbert spaces. Finally, we could derive optimal convergence rates for the IRGNM of inverse problems with Gaussian white noise based on the deterministic convergence results and the deviation inequality for Gaussian white noise derived by Veraar [92].

Chapter 7 is devoted to a generalization of the iteratively regularized Gauss-Newton method, where we analyze convergence results of (1.7) with Poisson and empirical process data. In order to improve rates of convergence (1.8) for Newton-type method, we still need assumptions on the nonlinearity of F and smoothness assumptions we stated in Chapter 3 and the large deviation inequalities derived in Chapter 4. Furthermore, we also recall once again the convergence results for iteratively regularized Newton-type method with Poisson and empirical process data studied by Hohage et al [51, 24]. Then we present our main achievements for the Newton-type method and derive convergence rates close to optimal order for Poisson and empirical process data.

Chapter 2

Inverse problems with stochastic data

In this chapter we will recall several classes of statistical noise models described by exponential families, where the distributions of the data for example are given by (Gaussian) white noise, Poisson or empirical processes. First of all, we will collect some basic properties of these processes for developing our analysis in inverse problems. In particular, inverse problems with Poisson data have been intensively studied over the last decades since they appear in many photonic imaging applications such as fluorescence microscopy, coherent x-ray imaging and positron emission tomography (see [9, 19, 51, 84]). Inverse problems with empirical process data arise in parameter identification problems in stochastic differential equations, which occur in financial econometrics, biology and geology [24, 77].

In section 2.1 we introduce statistical noise models that the measurements are perturbed by deterministic and stochastic random noises. Furthermore, we recall some required properties of these statistical models in inverse problems. In section 2.2 we will formulate the observed data drawn from a Poisson or empirical processes in the context of inverse problems respectively and present some essential results on concentration inequalities for Poisson processes and empirical processes. In section 2.3 some prominent examples of inverse problems with Poisson and empirical process data and applications will be discussed.

2.1   Additive random noise

We consider at first the deterministic error model where the observed data g^{obs} satisfy

$$g^{\text{obs}} = F(f^\dagger) + \xi, \quad \|\xi\| \leq \delta \tag{2.1}$$

with a deterministic noise level $\delta > 0$, $F : \mathbb{X} \to \mathbb{Y}$ is a forward operator between Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$ and $\xi \in \mathbb{Y}$ is deterministic error. The deterministic error model (2.1) and regularization methods in numerical analysis of inverse problems has been extensively explored in the literature of Hanke and Neubauer [28].

Recently, the investigation of inverse problems has also become of importance from a statistical point of view. For instance the measured data g^{obs} in many practical applications is corrupted by a stochastic random noise W on Hilbert space $\mathbb{Y}$ as following

$$g^{\text{obs}} = F(f^\dagger) + \varepsilon W \tag{2.2}$$

with the stochastic noise level $\varepsilon > 0$. In signal processing, white noise is a random signal having equal intensity at different frequencies and giving it a constant power

spectral density. The process appears as a generic limit process under mild assumptions. The stochastic error W satisfying

$$\mathbf{E}\langle g, W\rangle = 0, \quad \mathbf{E}\left[\langle g_1, W\rangle\langle g_2, W\rangle\right] = \langle g_1, g_2\rangle$$

for all $g \in \mathbb{Y}$ and $g_1, g_2 \in \mathbb{Y}$.

In the context of white noise model the white noise process $W : \mathbb{Y} \to L^2(\Omega, \mathcal{F}, \mathbf{P})$ is a Hilbert space process on $\mathbb{Y}$ with the underlying probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Any Hilbert space valued random variable Σ with finite second moment $\mathbf{E}\|\Sigma\|^2 < \infty$ can be identified with a Hilbert-space process given by the following map:

$$g \mapsto \langle \Sigma, g\rangle, \quad \text{for } g \in \mathbb{Y}.$$

Note that if W is a Gaussian white noise process in an infinite dimensional Hilbert space, i.e., $\langle W, g\rangle \sim N(0, \|g\|^2)$. Then $\|W\|$ is not finite almost surely and it cannot be identified with a Hilbert space valued random variable, i.e., $\mathbf{E}\|W\|_{\ell^2}^2 = \infty$. For more details, we refer to Pricop and Hohage [48].

As we introduced in Chapter 1, the negative log-likelihood used as data fidelity term in Bayesian statistics and it has good properties from a frequentist perspective. For the deterministic noise model (2.1), the choice of the data fidelity functional is given by

$$\mathcal{S}(g^{\text{obs}}, g) = \frac{1}{2}\|g^{\text{obs}} - g\|_{\mathbb{Y}}^2.$$

In contrast to the noise model (2.1), for statistical noise models (2.2) with Gaussian processes W one may try to defined the data fidelity functional by

$$\mathcal{S}(g^{\text{obs}}, g) = \frac{1}{2}\|g^{\text{obs}} - g\|_{\mathbb{Y}}^2 = \frac{1}{2}\|g^{\text{obs}}\|^2 - \langle g^{\text{obs}}, g\rangle + \frac{1}{2}\|g\|^2.$$

However, $\|g^{\text{obs}}\|^2$ is infinite almost surely in an infinite dimensional Hilbert space $\mathbb{Y}$ and $\mathcal{S}(g^{\text{obs}}, g) \equiv \infty$ is not a useful data fidelity term. Therefore, we subtract the infinite term $\|g^{\text{obs}}\|^2$, which does not depend on g such that we can use the following data fidelity term

$$\mathcal{S}(g^{\text{obs}}, g) := \frac{1}{2}\|g\|^2 - \langle g^{\text{obs}}, g\rangle. \tag{2.3}$$

For the exact data $g^{\dagger}$, we use

$$\mathcal{T}(g^{\dagger}, g) = \frac{1}{2}\|g^{\dagger} - g\|_{\mathbb{Y}}^2. \tag{2.4}$$

The functional $\mathcal{T}$ is a natural distance measure in the Hilbert space $\mathbb{Y}$ and non-negative. But the exact data fidelity functional $\mathcal{T}$ is same for the both noise models (2.1) and (2.2).

In the statistical convergence analysis of inverse problems, concentration or deviation inequalities play an essential role to study consistency and convergence rates for many estimators as well as they can apply for statistical inverse problems. We now restate an existing result of a concentration inequality for Gaussian process $\|W\|$, which describes the concentration of $\|W\|$ around its mean measured in terms of variance ϑ.

Lemma 2.1 (Concentration inequality for Gaussian process). *Let W be a Gaussian random variable in Banach space $\mathbb{X}$ with median M, let $\vartheta = \vartheta(W)$ be a supremum of weak*

variance. Then the following inequality holds true

$$\mathbf{P}\left(|\|W\| - M| > r\right) \leq \exp\left(-\frac{r^2}{2\vartheta^2}\right) \tag{2.5}$$

for every $r > 0$.

Proof. See Lemma 3.1 [62]. □

Remark 2.2. *This type of concentration inequality is needed to derive a deviation inequality for Gaussian white noise on the tail behavior of $\|W\|$ in some function spaces. For more details about the deviation inequality, see Chapter 4.*

2.2 Poisson and empirical process data

2.2.1 Empirical processes

In probability theory, an empirical process is also a stochastic process that describes the finite or infinite sequence of independent random variables and it occur 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 $g^\dagger$, then the problem can be formulated as a (possibly non-linear) ill-posed operator equation as (1.1). For the given observations $Y_j = y_j$, we can describe these observation by the empirical measure

$$G_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{Y_j}, \tag{2.6}$$

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

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

where $\mathbb{I}_{\mathbb{A}}$ is the indicator function of the set $\mathbb{A}$. Since the empirical process $G_n = \sum_{j=1}^{n}\delta_{y_j}$ can be seen as a random measure, we can define integrals over complex continuous function $\varphi : \mathbb{M} \to \mathbb{R}$ w.r.t. G_n, i.e.,

$$\int_{\mathbb{M}}\varphi \mathrm{d}G_n = \frac{1}{n}\sum_{j=1}^{n}\varphi(y_j).$$

For any complex continuous function $\varphi : \mathbb{M} \to \mathbb{C}$ with respect to G_n, we have

$$\mathbf{E}\left[\int_{\mathbb{M}}\varphi \mathrm{d}G_n\right] = \int_{M}g^\dagger\varphi \mathrm{d}x, \quad \mathbf{Var}\left[\int_{\mathbb{M}}\varphi \mathrm{d}G_n\right] = \frac{1}{n}\int_{M}g^\dagger|\varphi|^2\mathrm{d}x \tag{2.7}$$

whenever the integral on the right hand sides exist.

Similarly, there also exist a concentration inequality for suprema of empirical processes. In probability theory, Talangrand's concentration inequalities in product spaces [86] were well-known, who proved the following theorem for the suprema of empirical processes:

Theorem 2.3 (Talagrand inequalities). *Let $X_1, \cdots, X_n$ be independent identically distributed random variables on some measurable space $(\mathbb{M}, \Omega)$, and $\mathcal{F}$ is a countable family*

of real-valued measurable functions on $(\mathbb{M}, \Omega)$ *such that* $\sup_{\varphi \in \mathcal{F}} \|\varphi\|_\infty < \infty$ *for every* $\varphi \in \mathcal{F}$. *Let*

$$Z = \sup_{\varphi \in \mathcal{F}} \sum_{j=1}^{n} \varphi(X_j) \quad and \quad v = \mathbf{E} \left[\sup_{\varphi \in \mathcal{F}} \sum_{j=1}^{n} \varphi^2(X_j) \right].$$

Then for $x > 0$, *the following inequalities holds true*

$$\mathbf{P}\left(|Z - \mathbf{E}[Z]| \geq x\right) \leq C_1 \exp\left(-\frac{x}{C_2 b} \ln\left(1 + \frac{xb}{v}\right) \right) \tag{2.8}$$

and

$$\mathbf{P}\left(|Z - \mathbf{E}[Z]| \geq x\right) \leq C_1 \exp\left(-\frac{x^2}{2c_1 v + c_2 b x} \right), \tag{2.9}$$

where C_1, C_2, c_1 *and* c_2 *are universal positive constants.*

Proof. See Theorem 4.1 [86]   □

The inequalities (2.8) and (2.9) can be seen as functional version of Bennett's and Bernstein's inequalities respectively for sums of independent and real-valued random variables. If only $\mathbf{P}\left(Z - \mathbf{E}[Z] \geq x\right)$ is bounded, then we call it **deviation inequality** rather than concentration inequality. Afterwards, Massart [65] improved the Talangrand's concentration inequalities and obtained new inequalities for the suprema of empirical processes with explicit constants based on the result by Talagrand [86]. Here we state this concentration inequality with our notation:

Lemma 2.4 (Concentration inequality for empirical process). *Let* G_n *be an empirical process with intensity* $g^\dagger$ *and let* $\mathcal{F} \subset L^\infty(\mathbb{M})$ *be a countable family of functions with* $\|\varphi\|_\infty \leq b$ *for all* $\varphi \in \mathcal{F}$. *Moreover, let*

$$Z := n \sup_{\varphi \in \mathcal{F}} \left| \int_{\mathbb{M}} \varphi(\mathrm{d}G_n - g^\dagger \mathrm{d}x) \right| \quad and \quad v_0 := n \sup_{\varphi \in \mathcal{F}} \int_{\mathbb{M}} \varphi^2 g^\dagger \mathrm{d}x.$$

Then the following inequality holds true

$$\mathbf{P}\left[Z \geq (1 + \varepsilon)\mathbf{E}[Z] + 2\sqrt{2v_0 \eta} + \mu(\varepsilon) b\eta \right] \leq \exp(-\eta) \tag{2.10}$$

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

Proof. See Theorem 3 in [65].   □

Note that the difference between the weak variance v in Theorem 2.3 and the wimpy variance v_0 in Lemma 2.4. Due to several reasons such as expressing the variance factor with wimpy variance v_0 rather than weak variance v, a contraction principle gives that v is greater than v_0. We now recall in the following the result of Massart [65] with our notation:

Corollary 2.5. *Let* $X_1, \cdots, X_n$ *be independent identically distributed random variables on some measurable space* $(\mathbb{M}, \Omega)$, *and* $\mathcal{F}$ *is a countable family of real-valued measurable functions on* $(\mathbb{M}, \Omega)$ *such that* $\sup_{\varphi \in \mathcal{F}} \|\varphi\|_\infty < \infty$ *for every* $\varphi \in \mathcal{F}$. *Then,*

$$\mathbf{E}\left[\sup_{\varphi \in \mathcal{F}} \sum_{j=1}^{n} \varphi^2(X_j) \right] \leq \sup_{\varphi \in \mathcal{F}} \mathbf{E}\left[\sum_{j=1}^{n} \varphi^2(X_j) \right] + 16\mathbf{E}[Z].$$

Remark 2.6. *Massart [65] also obtained a similar concentration inequality for the left tail of Z, but we do not need this deviation inequality in our thesis.*

For the empirical process G_n, the probability density is given by

$$\mathbf{P}_g[y_1, \cdots, y_n] = \prod_{j=1}^{n} g(y_j)$$

and the corresponding negative log-likelihhod functional is given by

$$\mathcal{S}(G_n, g) = -\frac{1}{n} \ln \mathbf{P}_g[y_1, \cdots, y_n] = -\int_{\mathbb{M}} \ln g \, \mathrm{d}G_n, \tag{2.11}$$

which is a data fidelity functional for random data G_n. By applying (2.7), we have

$$\mathbf{E}\left[\mathcal{S}(G_n, g)\right] = \int_{\mathbb{M}} g^{\dagger} \ln(g) \, \mathrm{d}x \quad \text{and} \quad \mathbf{Var}\left[\mathcal{S}(G_n, g)\right] = \frac{1}{n} \int_{\mathbb{M}} \ln(g)^2 g^{\dagger} \, \mathrm{d}x.$$

The integral on the right hand side of (2.11) can be written as a sum of functional

$$\int_{\mathbb{M}} \ln g \, \mathrm{d}G_n = \sum_{j=1}^{N} \ln(g(x_j)),$$

which is infinite if $g(x_j) = 0$ for any $j \in \{1, \cdots, N\}$. It can be seen from (2.11) that the minimum value of $\mathcal{S}(G_n, g)$ has no significance since it is not symmetric and it does not necessarily satisfy a triangle inequality even though the minimum of $\mathcal{S}(G_n, g)$ is attained close to the true data $g^{\dagger}$. A non-negative distance between g and $g^{\dagger}$ can be obtained by taking expectations:

$$\mathbf{E}\left[\mathcal{S}(G_n, g) - \mathcal{S}(G_n, g^{\dagger})\right] = \begin{cases} \int_{\mathbb{M}} \left[g - g^{\dagger} + g^{\dagger} \ln\left(\frac{g^{\dagger}}{g}\right)\right] \mathrm{d}x, & \text{if } g \geq 0 \text{ a.e.,} \\ \infty & \text{else,} \end{cases} \tag{2.12}$$

where the integral on the right hand side of (2.12) is known as **Kullback-Leibler divergence** and denoted by $KL(g, g^{\dagger})$. Note that $KL(g, g^{\dagger})$ is a strictly convex and lower semicontinuous functional on $L^1(\mathbb{M})$ and $KL(g, g^{\dagger}) = 0$ if and only if $g = g^{\dagger}$ (see [26]). The Kullback-Leibler divergence is not only suitable forempirical process data but also connected to other types of exponential families [90].

Unfortunately, the data fidelity term (2.11) does not satisfy the condition of the concentration results in Lemma 2.4 since the logarithmic function $\ln(g)$ will be unbounded and g in many applications may be 0 in parts of the domain or get arbitrarily close to zero. For these reasons, we will introduce a modified data fidelity term $\mathcal{S}_{\zeta}(G_n, g)$ with a shift parameter $\zeta \geq 0$:

$$\mathcal{S}_{\zeta}(G_n, g) := \begin{cases} \int_{\mathbb{M}} g \, \mathrm{d}x - \int_{\mathbb{M}} \ln(g + \zeta)(\mathrm{d}G_n + \zeta \mathrm{d}x) & \text{if } g \geq -\frac{\zeta}{2} \text{ a.e.,} \\ \infty, & \text{else} \end{cases} \tag{2.13}$$

and

$$\begin{aligned} \mathcal{T}_{\zeta}(g^{\dagger}, g) &:= \mathbf{E}\left[\mathcal{S}_{\zeta}(G_n, g) - \mathcal{S}_{\zeta}(G_n, g^{\dagger})\right] \\ &= \begin{cases} KL(g^{\dagger} + \zeta, g + \zeta), & \text{if } g \geq -\frac{\zeta}{2} \text{ a.e.,} \\ \infty, & \text{else.} \end{cases} \end{aligned} \tag{2.14}$$

Note that the data fidelity functional $\mathcal{S}_\zeta$ is convex in its first argument and it is also true for the data fidelity functional in the case of noisy data $\mathcal{S}(G_n, \cdot)$.

It is important to emphasize the significance of the offset parameter $\zeta > 0$ due to the concentration inequality (2.10). If we had a concentration inequality for unbounded functions for example in L^2, then it is also possible to work for the case $\zeta = 0$. One of the reasons for choosing a small shift parameter $\zeta > 0$ is the instabilities of estimators and slow convergence of regularization methods for the case $\zeta = 0$. It is also pointed out by Hohage and Werner [53] that the choice of positive shift parameter $\zeta > 0$ is a common idea in many literatures on inverse problems with Poisson or empirical process data. Moreover, the numerical computations become easier for $\zeta > 0$ due to the singularity at 0.

2.2.2   Poisson point processes

In this subsection we will give a definition of a Poisson point process (or Poisson process) and collect some basic properties of Poisson processes. As a main reference on Poisson point processes we refer to [59].

Let $\mathbb{M} \subset \mathbb{R}^d$ be a Riemannian observation manifold such that a point process on $\mathbb{M}$ can be described 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 number of observed points N are random. In this setup, the measurement manifold $\mathbb{M}$ maybe an open set or sphere. Alternatively, a Poisson point process can be defined by the notion of a sum of Dirac measures. For the given observation $X_j = x_j$ on $\mathbb{M}$, the sum of Dirac-measures can be defined by

$$G = \sum_{j=1}^{N} \delta_{x_j}, \quad \text{for } j = 1, 2 \cdots, N \tag{2.15}$$

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 $\mathbb{A} \subset \mathbb{M}$. The Poisson process (2.15) could give us to count two or more points at the same position $x_j = x_{\tilde{j}}$ with $j \neq \tilde{j}$. However, we consider the Poisson process G as a set of random points in our definition such that all points x_j need to be pairwise different in our setting.

As described by Kingman [59], G is a Poisson process and $G(\mathbb{A})$ is a Poisson distributed integer-valued random variable, which can be clarified in the following definition:

Definition 2.7 (Poisson point processes). *Let $\mathbb{M}$ be a measurable set with intensity $g^\dagger \in L^2(\mathbb{M}) \geq 0$ for $g^\dagger \geq 0$. The Dirac measure (2.15) is called Poisson point process (or Poisson process) if the following properties are satisfied:*

- *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.*

- *For any measurable subset $\mathbb{A} \subset \mathbb{M}$, the mean measure $\lambda = \mathbf{E}[G(\mathbb{M})]$ exists and it is given by $\mathbf{E}[G(\mathbb{A})] = \int_\mathbb{A} g^\dagger \mathrm{d}x$.*

For more information about Poisson point processes, we refer to [55, 59]. Note that if $g^\dagger$ is normalized as a density function for an exposure time $n > 0$ and $\int_\mathbb{M} g^\dagger \mathrm{d}x = 1$, then n is equal to the expected total number of points, i.e., $n = \mathbf{E}[N]$.

One can show that for each measurable subset $\mathbb{A} \in \mathbb{M}$ the random variable $G(\mathbb{A})$ is Poisson distributed with parameter $\lambda := \int_{\mathbb{A}} g^{\dagger} \mathrm{d}x$, 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 [59]. 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 $\mathscr{P}(\mathbb{A})$ is give by

$$\mathscr{P}(\mathbb{A}) := \frac{\int_{\mathbb{A}} g^{\dagger} \mathrm{d}x}{\int_{\mathbb{M}} g^{\dagger} \mathrm{d}x}$$

and the distribution

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

For the proof, we refer to Proposition 2.2 in Hohage and Werner [53].

Since the Poisson process $G = \sum_{j=1}^{N} \delta_{x_j}$ can be seen as a random measure, we can again define integrals over complex continuous function $\varphi : \mathbb{M} \to \mathbb{R}$ with respect to G. i.e.,

$$\int_{\mathbb{M}} \varphi \mathrm{d}G = \sum_{j=1}^{N} \varphi(x_j).$$

For any complex continuous function $\varphi : \mathbb{M} \to \mathbb{C}$ with respect to G, we have

$$\mathbf{E}\left[\int_{\mathbb{M}} \varphi \mathrm{d}G\right] = \int_{M} g^{\dagger} \varphi \mathrm{d}x, \quad \mathbf{Var}\left[\int_{\mathbb{M}} \varphi \mathrm{d}G\right] = \int_{M} g^{\dagger} |\varphi|^2 \mathrm{d}x \qquad (2.16)$$

whenever the integral on the right hand sides exist (see [59]).

Definition 2.8. *Let G be a Poisson point process on $\mathbb{M}$ with mean λ, where $\mathbb{M}$ is a subset of $\mathbb{R}^d$. If the mean λ is absolutely continuous w.r.t. the Lebesgue measure, then the Radon-Nikodym derivative $g \in L^1(\mathbb{M})$ of λ w.r.t. the Lebesgue measure is called the intensity of G. If G has a constant intensity g, then G is called a homogeneous Poisson process.*

As introduced in the previous chapter, we need an additional parameter $n > 0$, which is called an **exposure time** (or observation time). That means the measurement manifold $\mathbb{M}$ to contain a time variable n with the intensity $ng^{\dagger}$ and we would like to consider a continuous measurement procedure for fixed $n > 0$, then the corresponding data will be a temporally rescaled Poisson process. Assume that the data $\widetilde{G}_n$ are drawn from a Poisson process with intensity $ng^{\dagger}$ and we define a rescaled Poisson process by $G_n = \frac{1}{n}\widetilde{G}_n$, where the scaling factor $\frac{1}{n}$ ensure that the expectation of the integral function is independent of $n > 0$, (see [53]). Thus, (2.16) yields

$$\mathbf{E}\left[\int_{\mathbb{M}} \varphi \mathrm{d}G_n\right] = \int_{M} g^{\dagger} \varphi \mathrm{d}x, \quad \mathbf{Var}\left[\int_{\mathbb{M}} \varphi \mathrm{d}G_n\right] = \frac{1}{n} \int_{M} g^{\dagger} \varphi^2 \mathrm{d}x. \qquad (2.17)$$

This indicates that the noise level should be determined by the scaling factor $\frac{1}{\sqrt{n}}$. Thus, we study the convergence rates and optimality for the convergence rates of estimator to inverse problem with the Poisson data G_n in the limit $n \to \infty$.

The known facts show that a normalized Poisson process $\frac{1}{n}\widetilde{G}_n$ concentrates more and more around its expectation $g^{\dagger}$ for $n \to \infty$ rather than the arbitrary detected

point process. For a Poisson process (2.6), it is even difficult to establish concentration inequalities due to the non-existence of symbolic expression for its expectation and variance. The general concentration inequality for a random variable is not useful for the type of supremum over all integrals of φ with respect to the Poisson process G_n. A uniform concentration inequalities for Poisson process has been derived by Reynaud-Bouret [76]:

Lemma 2.9 (Concentration inequality for Poisson process). *Let G_n be a Poisson process with intensity $g^\dagger$ and let $\mathcal{F} \subset L^\infty(\mathbb{M})$ be a countable family of functions with $\|\varphi\|_\infty \leq b$ for all $\varphi \in \mathcal{F}$. Moreover, let*

$$Z := n \sup_{\varphi \in \mathcal{F}} \left| \int_{\mathbb{M}} \varphi(\mathrm{d}G_n - g^\dagger \mathrm{d}x) \right| \quad and \quad v_0 := n \sup_{\varphi \in \mathcal{F}} \int_{\mathbb{M}} \varphi^2 g^\dagger \mathrm{d}x,$$

then the following inequality holds true

$$\mathbf{P}\left[Z \geq (1+\varepsilon)\mathbf{E}[Z] + 2\sqrt{3v_0\eta} + \mu(\varepsilon)b\eta \right] \leq \exp(-\eta) \tag{2.18}$$

for $\eta, \varepsilon > 0$, where $\mu(\varepsilon) = \frac{5}{4} + \frac{32}{\varepsilon}$.

Proof. See Corollary 2 [76]. $\qquad\square$

This type of concentration inequality is also established based on the Talagrand's concentration inequalities for the suprema of empirical processes and it has a quite similar form as Massart's concentration inequalities [65] with different constants. The negative log-likelihood functional $\mathcal{S}$ for Poisson process is slightly different from the empirical process. For the Poisson process, the data fidelity functional is derived by

$$\mathcal{S}(G_n, g) := -\ln \mathbf{P}_g[G_n] = \begin{cases} \int_{\mathbb{M}} g\mathrm{d}x - \int_{\mathbb{M}} \ln(g)\mathrm{d}G_n & \text{if } g \geq -\frac{\zeta}{2} \ a.e., \\ \infty, & \text{else.} \end{cases} \tag{2.19}$$

By using (2.16), we have

$$\mathbf{E}\left[\mathcal{S}(G_n, g)\right] = \int_{\mathbb{M}} g^\dagger \left(g - g^\dagger \ln(g) \right) \mathrm{d}x \quad \text{and} \quad \mathbf{Var}\left[\mathcal{S}(G_n, g)\right] = \frac{1}{n} \int_{\mathbb{M}} \ln(g)^2 g^\dagger \mathrm{d}x.$$

Again, the data fidelity functional (2.19) is difficult to deal with due to the fact that the logarithmic function $\ln(g)$ will not be bounded. Therefore, we will need a modified version of data fidelity functional with a shift parameter $\zeta \geq 0$. After simple computation, we have for

$$\mathcal{S}_\zeta(G_n, g) := \begin{cases} \int_{\mathbb{M}} g\mathrm{d}x - \int_{\mathbb{M}} \ln(g+\zeta)(\mathrm{d}G_n + \zeta\mathrm{d}x) & \text{if } g \geq -\frac{\zeta}{2} \ a.e., \\ \infty, & \text{else} \end{cases} \tag{2.20}$$

that

$$\begin{aligned} \mathcal{T}_\zeta(g^\dagger, g) := \ & \mathbf{E}\left[\mathcal{S}_\zeta(G_n, g) - \mathcal{S}_\zeta(G_n, g^\dagger) \right] \\ = \ & \begin{cases} \mathrm{KL}(g^\dagger + \zeta, g + \zeta), & \text{if } g \geq -\frac{\zeta}{2} \ a.e., \\ \infty, & \text{else.} \end{cases} \end{aligned} \tag{2.21}$$

To achieve the convergence rates for ill-posed problems with these stochastic processes, we need the following general assumption on the operator F.

Assumption 2.10 (Assumptions on forward operator *F*). *Let $\mathbb{X}$ be a Hilbert space and $\mathcal{B} \subset \mathbb{X}$ a bounded, closed and convex subset containing the exact solution $f^\dagger \in \mathcal{B}$. Let $\mathbb{Y} = L^2(\mathbb{M})$ for some bounded observation manifold $\mathbb{M} \subset \mathbb{R}^d$ with Lipschitz boundary $\partial\mathbb{M}$. Moreover, assume that the operator $F : \mathcal{B} \to \mathbb{Y}$ has the following properties:*

(i) It holds $F(f) \geq 0$ a.e. for all $f \in \mathcal{B}$.

(ii) F is injective operator on $\mathcal{B}$ and $f^\dagger \in \mathcal{B}$ is an exact solution of the exact problem $F(f) = g^\dagger$.

(iii) $F : \mathcal{B} \to \mathbb{Y}$ is continuously Fréchet differentiable with derivative $F'[f] : \mathbb{X} \to \mathbb{Y}$ for all $f \in \mathcal{B}$ with respect to L^2-norm.

Let us discuss these assumptions with the item (i): the property (i) is required due to the non-negativity of densities of the Poisson or empirical processes data. For linear operator *F* (i) is verified by proving that *F* preserves non-negativity and $\mathcal{B}$ is closed as a subset of the cone of non-negativity functions. The item (ii) is a natural assumption for inverse problems and the item (iii) is usually not difficult to verify in applications. Since the subset $\mathcal{B}$ is assumed to be closed and bounde, we define the finite quantity of the diameter of $\mathcal{B}$ by

$$\mathrm{diam}(\mathcal{B}) := \sup_{f, \tilde{f} \in \mathcal{B}} \|f - \tilde{f}\|_{\mathbb{X}},$$

which will be used for our convergence analysis in Chapter 5.

2.3   Some examples of inverse problems with stochastic data

In this section we will illustrate some well-known examples of inverse problems with stochastic data. In particular, the examples of inverse problems with Poisson data have been explored by many authors [19, 40, 51, 53, 54, 91] due to their relevance in photonic imaging from practical applications. In addition, empirical process data described by stochastic differential equations also applied in financial econometrics and physics (see e.g., [12, 20, 24, 60]). Dunker and Hohage [24] study in particular the convergence rates for parameter identifidication problems in stochastic differential equations.

In the photonic imaging, the Poisson distributed data consist of photon counts, where the detected photons interact with some unknown object $f^\dagger$. Such kind of photon interactions with unknown object $f^\dagger$ can be formulated by an operator equation $F(f^\dagger) = g^\dagger$, where the photon density $g^\dagger$ is a non-negative, integrable function on the measurement manifold $\mathbb{M}$ and *F* is a (possibly) nonlinear operator describes the imaging setup.

From the mathematical point of view, the photon counts can be seen as a random set of detected points such that the collected data can be described by (2.15) and satisfy the definition 2.7. In the following we will give an overview for some important examples from photonic imaging including inverse obstacle scattering problems without phase and fluorescence microscopy.

2.3.1   Phaseless inverse obstacle scattering problems

The identification of properties of inaccessible objects from measurements of scattered waves is called inverse scattering problem. Such problems arise for example in acoustics, time-harmonic electromagnetics and elasticity. The transverse

magnetic time harmonic electromagnetic waves is described by the Helmholtz equation with homogeneous Neumann boundary conditions

$$\Delta E + k^2 E = 0, \quad \text{in } \mathbb{R}^2 \setminus D, \tag{2.22a}$$

$$\frac{\partial E}{\partial \mathbf{n}} = 0, \quad \text{on } \partial D, \tag{2.22b}$$

$$\lim_{|x| \to \infty} \sqrt{|x|} \left(\frac{E_s}{|x|} - ikE_s \right) = 0 \tag{2.22c}$$

for the total field E. Here E is the sum of a known incident field E_i and a scattered field E_s, i.e., $E = E_i + E_s$, D is compact cross-section of the cylinder, $\mathbf{n}$ is the outer normal vector on boundary of D. In this class of obstacle scattering problem, the incident field $E_i(x) = \exp(ik\theta \cdot x)$ is considered to be a wave with direction $\theta \in \{x \in \mathbb{R}^2 | \ |x| = 1\}$. The scattered field E_s satisfies the Sommerfeld radiation condition (2.22c), which implies that the scattered field behaves asypmtotically like an outgoing wave

$$E_s(x) = \frac{\exp(ik|x|)}{\sqrt{|x|}} \left(E_\infty \left(\frac{x}{|x|} \right) + \mathcal{O} \left(\frac{1}{|x|} \right) \right)$$

as $|x| \to \infty$. The function E_∞ is defined on the sphere $S^1 = \{x/|x| \in \mathbb{R}^2 : x = |x|\}$ is called the far field pattern or scattering amplitude of E_s. In the classical inverse scattering problems, we need to find the obstacle D under some a-priori information and given measurements of E_∞ on the measurement manifold $\mathbb{M} = S^1$.

The aim of the inverse obstacle problem is to recover the shape of the obstacle D from photon counts in the field of scattered electromagnetics Since the photon density is proportional to the squared absolute value of the electric field, we are not able to immediately access to the phase of the electromagnetic field. Then the inverse problem is described by the operator equation

$$F(\partial D) = |E_\infty|^2,$$

where F is nonlinear. Moreover, for low energies the quantization of energy becomes important leading to photon count data with a shot noise as dominant source of noise. Then the data can be described by a Poisson process with density $|E_\infty|^2$ (see e.g. [51, 53, 54]). It can be shown that the far field pattern $|E_\infty|$ is invariant under translations of the boundary of D. Such inverse problems with other noise models have also been studied with different approaches by Ivanyshyn [54].

2.3.2   Fluorescence microscopy

One of the main applications of the inverse problems with Poisson data is confocal fluorescence microscopy. Here the fluorescent photons emitted in a small neighbourhood of the point are recorded by a detector. Let $p(x - y)$ denote the probability density of detecting a photon at point y for the focal point x, which depends on the special structure of the microscopy. Let f denote the fluorescent marker density. To recover f from the observed data, we can formulate the inverse problems with Poisson data through a linear operator equation

$$F(f)(x) = \int p(x - y) f(y) \mathrm{d}y,$$

where the density p is called *point spread function* and the density f fulfills the non-negativity assumption. Such problems also occur for example in astronomical imaging. For more details, we refer to Bertero et al [9]. Since the axial resolution of the standard confocal microscope cannot be achieved with the limited width of p, 4pi microscopy uses interference of coherent photons and the axial resolution be improved by a factor of 5-7.

Due to some experimental difficulties the relative phase of the interfering photons cannot always be determined, the aforementioned problems are not applicable for the image processing. However, we can assume that the phase ψ is a slowly varying function of x rather than a constant and it should be recovered with the unknown object f simultaneously. Thus, the imaging process can be described by the nonlinear forward operator

$$(F(f, \psi))(x) = \int p(x - y, \psi(x))f(y)\mathrm{d}y,$$

which is called a semiblind deconvolution problem due to the spread point function p depends on the unknown phase ψ. As discussed in Stück, Burger and Hohage [84], the point spread function is given by

$$p(x, \psi) = p(x) \cos^n \left(cx_3 + \frac{\psi}{2} \right),$$

where $h(x)$ is the point spread function of the confocal microscope and $n \in \{2, 4\}$ depends on the type of 4Pi-microscopy (see [39]). For more information on the fast implementation of the forward operator $F(f, \psi)$ and applications, we refer to Stück et al [84]. For the fast implementation of 4Pi-microscopy with the spread function $p(x, \psi)$ we refer to Stück et al [84], where they implement this microscopy with the spread function $p(x, \psi)$ by using the fast Fourier transform.

2.3.3 Parameter identification in stochastic differential equations

In this subsection we mainly focus on a parameter identification problem in stochastic differential equations. This type of example is motivated by the fact that empirical processes arise in financial econometrics and parameter identification problems, which can be described by Fokker-Planck equation

$$\mathrm{d}X_t = \mu(X_t, t)\mathrm{d}t + \sigma(X_t, t)\mathrm{d}W_t, \tag{2.23}$$

where $X_t = (X_{1,t}, \cdots, X_{d,t})'$ is a d-dimensional random variable with values in $\mathbb{R}^d$ and $t \in [0, \bar{t}]$ with $\bar{t} > 0$ can be interpreted as a time and W_t is standard Brownian motion in $\mathbb{R}^d$. The function $\mu : [0, \bar{t}] \times \mathbb{R}^d \to \mathbb{R}^d$ is a drift coefficient and $\sigma : [0, \bar{t}] \times \mathbb{R}^d \to \mathbb{R}^{d \times d}$ is the volatility (or diffusion). The equation system (2.23) describes dynamic processes in statistical mechanics such as physic, economics and other social sciences. Here we mainly focus on time homogenous differential equations that μ and σ are independent of time t and σ is known while μ should be estimated. For further details on parameter identification problems described by a partial differential equation, we refer to [5].

The modeling process (2.23) became popular in financial econometrics [20, 41, 60, 72, 74, 82]. Especially, the parametric and nonparametric estimation for the drift and diffusion in ergodic models are mostly investigated as an interest. Recently, Dunker and Hohage [24] consider the first time ill-posedness of such problems and studied

general convergence rates for variational regularization in terms of some suitable index function. There exist two different kinds of observations:

(i) A collection of independent paths $X_{j,t}$ for $j = 1, \cdots, n$ are observed at a fixed time $t = \bar{t}$, then the observation data are random variables, i.e., $Y_j = X_{j,\bar{t}}$. The initial values $X_{j,0}$ are taken from a known distribution g_0.

(ii) Only one path of a strictly stationary and ergodic process at equidistant times are observed for $Y_j = X_{(j+j_0)\Delta t}$ with $j_0 > 0$.

Note that the equidistant observations $Y_j = X_{(j+j_0)\Delta t}$ of one path are not independent, however, we can ignore this case while the information order is not available. The method to the problem can be established based on the Fokker-Planck equation (or forward Kolmogorov equation). We recall some basic properties of the stationary Fokker-Planck eqution and consider this equation on a bounded Lipschitz domain $\mathbb{M} \in \mathbb{R}^d$ with the non-flux boundary condition. For the Fokker-Planck equation, the natural boundary conditions are given by

$$
\begin{aligned}
&-\frac{\partial}{\partial x}\mu g + 2^{-1}\sigma\sigma^T \nabla g = 0, \quad \text{in } \mathbb{M} \\
&-g(\mu \cdot n) + 2^{-1}\left(\sigma\sigma^T \nabla g\right) \cdot = 0, \quad \text{on } \partial\mathbb{M} \\
&\int_{\mathbb{M}} g(x)\mathrm{d}x = 1.
\end{aligned} \tag{2.24}
$$

We assume that the drift coefficient μ and the volatility σ are well-defined in L^∞ and traces on the boundary condition $\partial\mathbb{M}$. There exists a constant $C_\sigma > 0$ such that

$$
C_\sigma|\gamma|_2 \leq |\sigma^T(x)\gamma|_2.
$$

for all $\gamma \in \mathbb{R}^d$ and all $x \in \bar{\mathbb{M}}$.

Now we check the different cases of dimension d on the boundary condition of the Fokker-Planck equation. For the case $d = 1$, without loss of generality we can assume that the Riemannian manifold is $\mathbb{M} = (-1, 1)$. The diffusion μ and the volatility σ can be extended by

$$
\mu(x) := \mu(1) \quad \mu(-x) := \mu(-1) \quad \text{for } x > 1
$$

and

$$
\sigma(x) := \sigma(1) \quad \sigma(-x) := \sigma(-1) \quad \text{for } x > 1,
$$

respectively. It can be seen from (2.24) that the constant coefficient equation

$$
-\mu g' + \frac{\sigma^2}{2}g'' = 0 \quad \text{with } \mu \neq 0
$$

has two independent solutions, which are 1 and $\exp(2\mu x/\sigma^2)$, and this function on real line has an integrable solution if and only if $\mu(1) < 0$ and $\mu(-1) > 0$. Thus, these solution satisfy the boundary condition in (2.24).

If X_t has a smooth density $g(\cdot, t)$ for all $t \in [0, \bar{t}]$, then (2.24) holds true with the solution of the initial value problem

$$
\begin{aligned}
&\frac{\partial g}{\partial t} = -\frac{\partial}{\partial x}\mu g + \frac{1}{2}\frac{\partial^2}{\partial x^2}\sigma\sigma^T g, \\
&g(x, 0) = g_0
\end{aligned} \tag{2.25}
$$

in [77]. If the stationary ergodic proceess with parameters μ and σ are independent from t such that the solutions to the equation (2.25) tend to a stationary solution as $t \to \infty$, which gives

$$0 = -\frac{\partial}{\partial x}\mu g + \frac{1}{2}\frac{\partial^2}{\partial x^2}\sigma\sigma^T g,$$
$$\int g(x)\mathrm{d}x = 1.$$

It is easy to find that the the solutions of boundary condition (2.24) are restricted on the domian $\mathbb{M} = (-1, 1)$. Here we can define the coefficient to solution (possibly nonlinear) operator by $F(\mu) := g$.

For higher dimensions of $d > 1$, the exact boundary conditions are nonlocal. However, we expect the solutions to (2.24) converges to a stationary solution for the Fokker-Planck equation in $\mathbb{R}^d$ as the domain $\mathbb{M}$ increase. Furthermore, in a biological aspect, diffusion σ in cells the solutions X_t are contained in $\mathbb{M} \subset \mathbb{R}^d$. In this sense, the solution paths X_t hits its boundary, which could be reflected in some way with a certain probability (see [80]).

In this modeling process, the unknown quantity of the inverse problems will be denoted by f, which can be a parameter in the model of μ such that $f = \mu$. We assume that $f^\dagger$ is the exact solution and $g^\dagger$ corresponds to the probability density of the independent random variables $Y_1, \cdots, Y_n$. Therefore, we can formulate this problems by an nonlinear operator equation $F(f^\dagger) = g^\dagger$. Furthermore, we can also describe the observation data by the empirical measure as in (2.6) for the given observations $Y_j = y_j$.

Chapter 3

Tikhonov and general spectral regularization with additive random noise

This chapter mainly deals with convergence rates for Tikhonov or variational regularization and general spectral regularization of mildly ill-posed inverse problems with additive random noise in Hilbert spaces, respectively. Inverse problems with additive random noise in a deterministic framework and in statistical context have been intensively studied in different aspects by many authors [10, 11, 17, 47, 48, 50, 93]. In the present thesis we especially study the convergence analysis for linear statistical inverse problems of variational regularization with additive Gaussian white noise in a statistical setup.

In order to make our thesis more readable, we would like to recall at first general spectral regularization methods, some regularization assumptions and well-known spectral source conditions as well as restate the main results on the optimal rates of convergence for general spectral regularization methods derived by Bissantz et al [11], all these important details are given in section 3.1. Furthermore, we study convergence results for regularization methods based on the variational analysis and introduce source conditions in the form of variational inequalities. We can also achieve optimal convergence rates for variational regularization under Hölder-type source conditions in certain situation in section 3.2. Finally, several substantial advantages of Tikhonov (variational) regularization methods compared to the spectral regularization will be discussed in section 3.2.3. Due to the specific structure of the problems with additive random noise in Hilbert spaces with quadratic penalties, these two tools are available for deriving the optimal convergence rates. However, the main difference in the convergence analysis of spectral and variational methods is that if the data fidelity functional $\mathcal{S}$ is not quadratic, then the general spectral theory is no longer applicable, which will be further studied in Chapter 4.

3.1 General spectral regularization

We will start this section with an overview over the statistical convergence analysis for general spectral regularization methods of linear inverse problems with additive random noise, where the operator is assumed to be linear, bounded and injective and denoted by $T : \mathbb{X} \to \mathbb{Y}$ throughout this chapter.

Over the last two decades, there exists a large amount of literature on regularization methods for linear and nonlinear inverse problems with additive random noise

[7, 10, 11, 48, 70, 71]. A number of publications mostly focused on methods that require the explicit knowledge of spectral decomposition, the simplest method is spectral cut-off or truncated singular value decomposition for compact operators (See, e.g., [25] and [64]). Another major approach would be wavelet-vaguelette methods and the functional form for an estimator are similar to a spectral cut-off estimator. However, here we only recall convergence results on the regularization methods based on spectral theory. General spectral regularization methods of linear inverse problems in both deterministic and stochastic setting and their applications have been extensively investigated by Bissantz et al [11]. One of the main difficulties in the analysis of general spectral regularization methods is the estimation of the integrated variance of the noise. In the statistical framework, the bound on the noise term depends not only on the regularization parameter but also on the distribution of the singular values of compact operator T. Therefore, some additional conditions on the operator are needed for analyzing our problems. For Tikhonov regularization and other methods, the spectral data of the operator is not required in the implementation. We briefly overview of regularization properties and source conditions in the following subsection.

3.1.1   Regularized estimators and smoothness assumptions

In spectral regularization theory, the spectral estimators of the exact solution $f^\dagger$ can be described by regularization methods of the form

$$\widehat{f}_\alpha = R_\alpha g^{\text{obs}}. \tag{3.1}$$

Here the observed data g^{obs} is given by either (2.1) or (2.2) and R_α is a bounded approximation to the Moore -Penrose inverse $T^\dagger$, i.e., $R_\alpha \approx T^\dagger$ and it is given by

$$R_\alpha = q_\alpha(T^*T)T^*, \tag{3.2}$$

where $q_\alpha \in \sigma(T^*T) \to \mathbb{R}$ is a collection of bounded filter function depend on some regularization parameter $\alpha > 0$. One of the special form of the regularized estimators $\widehat{f}_\alpha$ is the spectral cut-off estimator, which is given by

$$q_\alpha^{sc}(\lambda) := \frac{1}{\lambda} \quad \text{for } \lambda \geq \alpha \quad \text{and} \quad q_\alpha^{sc}(\lambda) = 0, \quad \text{for } \lambda < \alpha.$$

In the following, we need a number of smoothness assumptions on the filter functions q_α satisfied for all commonly used regularization methods including Tikhonov regularization, iterated Tikhonov regularization, Lardy's method, Landweber iteration and ν-methods. For more details of these methods, we refer to Engl et al [28].

Assumption 3.1 (Regularity assumption). *Let there exist some constants $C_{q,1}, C_{q,2} > 0$ such that the function defined by (3.2) satisfy the following three statements for the parameter $\alpha \in \mathcal{A} \subset (0, \infty)$:*

(i) *$\sup_\lambda |\lambda q_\alpha(\lambda)| \leq C_{q,1}$ for all $\lambda \in [0, \|T^*T\|]$.*

(ii) *$\sup_\alpha \sup_\lambda |\alpha q_\alpha(\lambda)| \leq C_{q,2}$ for all $\lambda \in [0, \|T^*T\|]$ and for all $\alpha \in \mathcal{A}$.*

(iii) *Let a residual function $r_\alpha(\lambda) = 1 - \lambda q_\alpha(\lambda)$ and there exists a constant $C_\nu > 0$ and a number $\nu_0 > 0$ called **qualification** of the mehtod such that $\sup_\alpha |\lambda^{\nu/2} r_\alpha(\lambda)| \leq C_\nu \alpha^{\nu/2}$ for all $\alpha \in \mathcal{A}$ and all $\nu \in (0, \nu_0]$.*

Definition 3.2 (Index function). *A function $\varphi : [0, A] \to [0, \infty)$ with $A > 0$ is called index function if it is continuous, strictly increasing and $\varphi(0) = 0$.*

For general source conditions, we will assume that there exists a constant C_φ such that

$$\sup_{\lambda \in \sigma(T^*T)} |\varphi(\lambda)(1 - \lambda q_\alpha(\lambda))| \le C_\varphi \varphi(\alpha), \quad \alpha \searrow 0. \tag{3.3}$$

This follows from condition (iii) under mild assumptions, see Mathé and Pereverzev [66]. Note that the condition (iii) implies condition (i) for $\nu = 0$ and the residual functions r_α is equivalent to

$$\lim_{\alpha \to 0} r_\alpha(\lambda) = \begin{cases} 0, & \lambda > 0, \\ 1, & \lambda = 0. \end{cases}$$

For Tikhonov regularization, we have

$$q_\alpha(\lambda) = \frac{1}{\lambda + \alpha} \quad \text{and} \quad r_\alpha(\lambda) = \frac{\alpha}{\lambda + \alpha}$$

for all $\alpha > 0$ with constants $C_\nu = 1$ and $C_{q,1} = 1$.

Note that the constants in Assumption 3.1 can be derived by computing filter functions q_α and residual functions r_α for different regularization methods including Tikhonov regularization, iterated Tikhonov regularization, Landweber iteration and ν-methods. For the computation of the constants with these methods, we refer to [50].

It is well-known that the convergence of any regularization method can be arbitrarily slow in general. Therefore, to derive suitable rates of convergence, we will measure the smoothness of the solution $f^\dagger$ related to the smoothness properties of T^*T in terms of some **source conditions**. Let φ be an index function and there exists a source $\omega \in \mathbb{X}$ such that

$$f^\dagger = \varphi(T^*T)\omega, \tag{3.4}$$

see [28, 29]. This type of condition is called **general spectral source condition**. In particular, the condition (3.4) can be reduced to an exact source condition $f^\dagger = T^*\omega$ for $\varphi(\lambda) = \sqrt{\lambda}$, see [28]. Furthermore, To derive the better rates of convergence, we are also interested in the source set of all $f^\dagger$ with $\|\omega\|_\mathbb{X} \le \rho$ for some $\rho > 0$, which will be denoted by

$$\mathcal{K}_{\varphi,\rho} := \{\varphi(T^*T)\omega : \omega \in \mathbb{X}, \|\omega\| \le \rho\}. \tag{3.5}$$

For mildly ill-posed inverse problems, a **Hölder-type source conditon** is usually appropriate for finitely smoothness operators T, which are given by

$$f^\dagger = (T^*T)^{\nu/2}\omega, \qquad \|\omega\| \le \rho. \tag{3.6}$$

Hölder-type source conditions apparently correspond to the choice $\varphi_\nu(\lambda) = \lambda^{\nu/2}$. The qualification of a method describes that can be explained by the method of the maximal degree of smoothness in terms of the Hölder-type source conditions (3.6).

The conditions (3.6) are suitable for example in the case of numerical differentiation. Define a Hilbert space

$$L^2_\diamond([0, 2\pi]) := \left\{ f \in L^2([0, 2\pi]) \Big| \int_0^{2\pi} f(x)\mathrm{d}x = 0 \right\}$$

of L^2 functions with zero mean. Then the inverse of the differentiation operator T on $L^2_\diamond([0, 2\pi])$ is given by

$$(Tf)(x) = \int_0^x f(z) \mathrm{d}z - \frac{1}{2\pi} \int_0^{2\pi} \int_0^x f(z) \mathrm{d}z \mathrm{d}x$$

for $x \in [0, 2\pi]$. If the solution $f^\dagger$ is an element of Sobolev space $H^{2\nu}_{per}([0, 2\pi])$, then we have $\varphi = \varphi_\nu$.

For exponentially ill-posed inverse problems, (3.6) is usually too restrictive such that an appropriate choice of λ is **logarithmic**:

$$\varphi_p(\lambda) := \begin{cases} (-\ln \lambda)^{-p}, & 0 < \lambda \leq \exp(-1) \\ 0, & \lambda = 0 \end{cases} \tag{3.7}$$

with $p > 0$, for such problems we imposed a logarithmic source condition

$$f^\dagger = \varphi_p(T^*T)\omega, \qquad \|\omega\| \leq \rho, \tag{3.8}$$

see Hohage [44, 46]. In order to avoid the singularity of $\varphi_p(\lambda)$ at $\lambda = 1$, we assume in the context that the norm in $\mathbb{X}$ is scaled such that

$$\|T^*T\| = \|T\|^2 \leq \exp(-1).$$

If singular values σ_j of the operator T decay to 0 at a polynomial rate, then (3.6) is more reasonable with $\varphi = \varphi_\nu$. If σ_j decays exponentially to 0, then the source conditions (3.7) is suitable with $\varphi = \varphi_p$ rather than $\varphi_\nu = \lambda^{\nu/2}$.

For the backwards heat equation, the logarithmic source conditions are suitable for determining the initial value $u(\cdot, 0)$ from the heat distribution $u(\cdot, \bar{t})$ at the late time $\bar{t} > 0$. We define the forward operator $T = c \exp(-\bar{t}\Delta)$ for some constant $c > 0$ with underlying spaces $\mathbb{X} = \mathbb{Y} = L^2(\mathbb{M})$, where Δ is the Laplace operator with same boundary conditions, see Hohage [46]. If we let the domain by $\mathbb{M}$ with periodic boundary conditions, then the range of $\varphi_p(T^*T)$ is exactly equal to $H^{2p}([0, 2\pi])$, i.e.,

$$\mathrm{Ran}(\varphi_p(T^*T)) = H^{2p}([0, 2\pi]),$$

Therefore, the range condition (3.4) with $\varphi = \varphi_p$ is equivalent to the true solution $f^\dagger$ in $H^{2p}([0, 2\pi])$.

3.1.2 Error measures and optimal convergence rates

In the statistical framework, the expected square error between the estimator $\widehat{f}_\alpha$ and the solution $f^\dagger$ is used as an error measure. For the white noise model (2.2), $\mathbf{E}[q_\alpha(T^*T)T^*W] = 0$. By the bias-variance decomposition, we have

$$\mathbf{E}\left[\|\widehat{f}_\alpha - f^\dagger\|_\mathbb{X}^2\right] = \left\|\mathbf{E}\left[\widehat{f}_\alpha\right] - f^\dagger\right\|_\mathbb{X}^2 + \varepsilon^2 \mathbf{E}\left[\|q_\alpha(T^*T)T^*W\|^2\right], \tag{3.9}$$

where the bias term $\|\mathbf{E}[\widehat{f}_\alpha] - f^\dagger\|$ can be estimated by standard methods (see Engel et al [28]). Under Assumption 3.1, we can bound the bias by

$$\left\|\mathbf{E}\left[\widehat{f}_\alpha\right] - f^\dagger\right\|_\mathbb{X} = \left\|(1 - q_\alpha(T^*T)T^*T)f^\dagger\right\| \leq C_\nu \varphi(\alpha)\rho + \frac{C_q}{\sqrt{\alpha}}\delta. \tag{3.10}$$

To bound the the variance term, we will apply a special method as described in Bissantz et al [11]. Assume that the operator T is compact and T^*T is a trace-class operator. Let $\{(u_j, v_j, \sigma_j) : j \in \mathbb{N}\}$ be a sigular value decomposition of T, where σ_j are sigular values and u_j, v_j are the orthonormal basis. Under Assumption 1 in Bissantz et al [11], we have

$$
\begin{aligned}
\varepsilon^2 \mathbf{E}\left[\|q_\alpha(T^*T)T^*W\|^2\right] &= \varepsilon^2 \text{trace}(\mathbf{Cov}[q_\alpha(T^*T)T^*W]) \\
&= \varepsilon^2 \sum_{j=1}^\infty q_\alpha(\sigma_j^2)^2 \sigma_j^2 \langle \mathbf{Cov}[W]v_j, v_j \rangle \\
&\leq \varepsilon^2 \sum_{j=1}^\infty q_\alpha(\sigma_j^2)^2 \sigma_j^2. \quad (3.11)
\end{aligned}
$$

Note that here we consider slightly different way for bounding the variance than Bissantz et al [11] did. Instead of thinking of the mean integrated variance with the integral with respect to Borel measure σ, we think of the variance bounded by the sum of filter functions q_α. We now consider a counting function to bound the variance rather than conditions on the asymptotic properties of sigular values σ_j. For $\alpha > 0$, we introduce

$$
\Gamma(\alpha) := \#\{j \in \mathbb{N} : \sigma_j^2 \geq \alpha\}. \quad (3.12)
$$

Here $\#U$ denotes the cardinality of a set U. Under Assumption 3.1, the sum on the right hand side of (3.11) can be written with the counting function Γ as follows

$$
\begin{aligned}
\sum_{j=1}^\infty q_\alpha(\sigma_j^2)^2 \sigma_j^2 &= -\int_0^\infty q_\alpha(\theta)^2 \theta d\Gamma(\theta) \\
&\leq \left(-\frac{C_{q,1}}{\alpha}\right)^2 \int_0^\alpha \theta d\Gamma(\theta) - C_{q,2}^2 \int_\alpha^\infty \frac{1}{\theta} d\Gamma(\theta). \quad (3.13)
\end{aligned}
$$

Since the counting function Γ is not smooth in general, we will assume that Γ can be approximated suitably by a smooth function $\widetilde{\Gamma}$ with properties similar to those Γ as $\alpha \searrow 0$. The function $\widetilde{\Gamma}$ is monotonically decreasing and it satisfies

$$
-\int_0^\infty \alpha d\Gamma(\alpha) < \infty \quad \text{and} \quad \lim_{\alpha \searrow 0} \alpha \Gamma(\alpha) = 0.
$$

Definition 3.3. *Let $\mathbb{X}, \mathbb{Y}$ be Hilbert spaces and $HS(\mathbb{X}, \mathbb{Y})$ is the set of all compact linear operators $T : \mathbb{X} \to \mathbb{Y}$ and the sequences of sigular values $\sigma_j(T) \in \ell^2(\mathbb{N}_0)$. The Hilbert-Schmidt norm is defined as*

$$
\|T\|_{HS}^2 = \text{Tr}(T^*T) := \left(\sum_{j=0}^\infty \sigma_j(T)^2\right)^{\frac{1}{2}}.
$$

*The elements of $HS(\mathbb{X}, \mathbb{Y})$ are called **Hilbert-Schmidt operators**.*

If T is Hilbert-Schmidt operator, then it follows from the fact that $-\int_0^\infty \alpha d\Gamma(\alpha) = \sum_{j=1}^\infty \sigma_j^2$. Finally, it follows from Lebesgue's dominated convergence theorem that

$$
\lim_{\alpha \searrow 0} \alpha \Gamma(\alpha) = \lim_{\alpha \searrow 0} \sum_{j=0}^\infty \alpha \mathbb{I}\{\sigma_j^2 \geq \alpha\} = 0.
$$

Assumption 3.4. *Let Γ be defined by (3.12).*

(i) *There exists a function $\widetilde{\Gamma} \in C^2((0,\infty))$ and constants $\bar{\alpha} \in (0, \|T^*T\|]$ and $h > 0$ such that*

$$\frac{1}{1+h}\widetilde{\Gamma}(\alpha) \leq \Gamma(\alpha) \leq (1+h)\widetilde{\Gamma}(\alpha), \quad \text{for all } \alpha \in (0, \bar{\alpha}], \tag{3.14a}$$

$$\lim_{\alpha \to \infty} \widetilde{\Gamma}(\alpha) = \lim_{\alpha \to \infty} \widetilde{\Gamma}'(\alpha) = 0, \tag{3.14b}$$

$$\widetilde{\Gamma}' < 0, \tag{3.14c}$$

$$-\alpha \widetilde{\Gamma}'(\alpha) \text{ is integrable on } (0, \bar{\alpha}], \tag{3.14d}$$

$$\lim_{\alpha \downarrow 0} \alpha \widetilde{\Gamma}(\alpha) = 0. \tag{3.14e}$$

(ii) *There exists a constant $K_{\widetilde{\Gamma}} \in (0, 2)$ such that for all $\alpha \in (0, \bar{\alpha}]$*

$$\frac{\widetilde{\Gamma}''(\alpha)}{-\widetilde{\Gamma}'(\alpha)} \leq \frac{K_\Gamma}{\alpha}. \tag{3.15}$$

(iii) *There exists a constant $C_{\widetilde{\Gamma}} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$*

$$\frac{C_{\widetilde{\Gamma}}}{\alpha} \leq \frac{-\widetilde{\Gamma}'(\alpha)}{\widetilde{\Gamma}(\alpha)}. \tag{3.16}$$

Here Condition (iii) will only be needed for the derivation of lower bounds and it is usually satisfied for mildly ill-posed problems but not for exponentially ill-posed problems, where $\widetilde{\Gamma}(\alpha) \sim c(-\ln \alpha)^p$ for constants $c, p > 0$.

Now we will present the convergence results of general spectral regularization methods for statistical linear inverse problems derived by Bissantz et al [11], where we use our notations.

Theorem 3.5 (Upper bounds for spectral regularization). *Suppose that Assumptions 3.1 and 3.4 (i)-(ii) hold true for the operator $T \in HS(\mathbb{X}, \mathbb{Y})$ and let $\{q_\alpha\}$ be a filter with qualification φ satisfying condition (3.3). If the general spectral estimator is given by $\widehat{f_\alpha} = q_\alpha(T^*T)T^*g^{obs}$, then the risk of $\widehat{f_\alpha}$ can be bounded by*

$$\mathbf{E}\left[\|\widehat{f_\alpha} - f^\dagger\|_{\mathbb{X}}^2\right] \leq \left(C_\varphi \varphi(\alpha)\rho + C_q \frac{\delta}{\sqrt{\alpha}}\right)^2 + C_\varepsilon^2 \frac{\widetilde{\Gamma}(\alpha)}{\alpha}\varepsilon^2 \tag{3.17}$$

for all $\alpha > 0$ with constant $C_\varepsilon = \sqrt{(2h + C_\varphi + 2hC_\varphi)C_{q,1}^2 + (1+h)C_{q,2}^2} > 0$, where $\kappa_{\widetilde{\Gamma}} = \frac{K_{\widetilde{\Gamma}}}{2 - K_{\widetilde{\Gamma}}}$. With the a-priori parameter choice rule

$$\alpha^* = \max\left(\Theta^{-1}\left(\frac{\delta}{\rho}\right), \left(\frac{\Theta}{\sqrt{\widetilde{\Gamma}}}\right)^{-1}\left(\frac{\varepsilon}{\rho}\right)\right), \tag{3.18}$$

where $\Theta(\alpha) := \sqrt{\alpha}\varphi(\alpha)$, we obtain

$$\sqrt{\mathbf{E}\left[\|\widehat{f_{\alpha^*}} - f^\dagger\|_{\mathbb{X}}^2\right]} \leq C_\rho \rho \varepsilon \left(\max\left(\Theta^{-1}\left(\frac{\delta}{\rho}\right), \left(\frac{\Theta}{\sqrt{\widetilde{\Gamma}}}\right)^{-1}\left(\frac{\varepsilon}{\rho}\right)\right)\right) \tag{3.19}$$

for sufficiently small $\varepsilon, \delta > 0$ as $\alpha \to 0$, where

$$C_\rho = (C_\varphi + C_q + C_\varepsilon) \quad and \quad C_q = \sqrt{C_{q,1} C_{q,2}} > 0.$$

Proof. See Theorem 3 [11], where we additionally use (3.13) to bound the propagated noise error. To bound (3.13) we will apply Lemma 12 and 13 and this yields (3.17). □

Note that we only need the error bound between $\widehat{f}_\alpha$ and $f^\dagger$ in Hilbert space $\mathbb{X}$, the additional convergence results in image space $\mathbb{Y}$ are considered as the second part of Theorem 3 in Bissantz et al [11]. Furthermore, we also present the convergence rates with a-*priori* parameter choice (3.18) and consider the sum of filter functions q_α and apply (3.13) instead of the continuous procedure as described by Bissantz et al [11]. The main difference between the deterministic and stochastic error analysis is the treatment of the noise error. For the statistical inverse problems, the estimation of the noise error and the convergence rates of the mean integrated square error does not depend only on the relative smoothness of the solution, but also on the distribution of the singular values of operator T. Now we will show the optimal convergence results for the general spectral regularization under Hölder-type source conditions.

Corollary 3.6. *Assume that the sigular values of T are given by $\sigma_j = j^{-p}$ for $j = 1, 2, \cdots$ with $p > 1/2$. Let Assumptions 3.1 and 3.4 hold true for the operator T. If*

$$\mathbf{E}\left[\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^2\right] \leq \left(C_\nu \alpha^\nu \rho + C_q \frac{\delta}{\sqrt{\alpha}}\right)^2 + C_\varepsilon \varepsilon^2 \alpha^{-1-\frac{1}{2p}} \tag{3.20}$$

for all $\alpha > 0$ with constant $C_\varepsilon > 0$. Then under Hölder source condition $\Gamma_\nu(\alpha) = \alpha^{\nu/2}$ with the a-priori parameter choice rule

$$\alpha^* = \max\left(\left(\frac{\delta}{\rho}\right)^{\frac{\nu}{\nu+1}}, \left(\frac{\varepsilon}{\rho}\right)^{\frac{2b\nu}{2b\nu+2b+1}}\right), \tag{3.21}$$

we obtain the following convergence rates

$$\sqrt{\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2\right]} \leq C\rho \max\left(\left(\frac{\delta}{\rho}\right)^{\frac{2\nu}{2\nu+1}}, \left(\frac{\varepsilon}{\rho}\right)^{\frac{4\nu}{4\nu+2+1/p}}\right) \tag{3.22}$$

for a constant $C = C_\nu + C_q + C_\varepsilon > 0$.

Proof. Since the sigular values of T are $\sigma_j = j^{-p}$ for $p > 1/2$, an upper bound on counting function $\Gamma(\alpha) = \#\{j \in \mathbb{N} : \sigma_j^2 \leq \alpha\}$ is given by

$$\widetilde{\Gamma}(\alpha) := \alpha^{-\frac{1}{2p}}$$

due to $\widetilde{\Gamma}(\alpha) = j = \Gamma(\alpha)$ for all $j \in \mathbb{N}$. It is easy to see that Assumption 3.4 holds true for this case. Under Hölder source condition $\Gamma_\nu(\alpha) = \alpha^{\nu/2}$, we have

$$\frac{\Theta(\alpha)}{\sqrt{\psi(\alpha)}} = \alpha^{\frac{2b\nu+2b+1}{4b}} \quad and \quad \Theta(\alpha) = \alpha^{\frac{\nu+1}{2}}.$$

Then Theorem 3.5 yields the rates as in (3.22). □

Example 3.7. (Backwards heat equation). We consider the periodic heat equation

$$\frac{\partial u}{\partial t}(x,t) = \Delta u(x,t), \quad u(x,0) = f(x)$$

for $x \in [0,a]^d, t \in (0,\bar{t})$ with the solution $u(\cdot,t) = \exp(t\Delta)f$. The forward operator $T : u(\cdot,0) \mapsto u(\cdot,\bar{t})$ is given by $T = \exp(-\bar{t}\Delta)$. For the exponentially ill-posed inverse problems, the source conditions are of the logarithmic form $\varphi_p(\lambda) = (-\ln t)^{-p}$ for $p > 0$, then the counting function is defined by

$$\Gamma(\alpha) := \#\{j \in \mathbb{N} : \sigma^2 \geq \alpha\} = \#\{\exp(-2\bar{t}\lambda_j) \geq \alpha\},$$

the choice of approximated function $\widetilde{\Gamma}(\alpha)$ is given by the form

$$\widetilde{\Gamma}(\alpha) = c \left(-\frac{1}{2\bar{t}} \ln \alpha \right)^{d/2}.$$

From Theorem 3.5, we obtain the following optimal rates

$$\sqrt{\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2 \right]} = O\left(\rho(-\ln \max(\delta,\varepsilon))^{-p} \right),$$

as $\max(\delta,\varepsilon) \to 0$.

3.1.3   Lower bounds

To show order optimality of the spectral estimators $q_\alpha(T^*T)T^*$, we will recall definitions of some risk estimators.

Definition 3.8. *Consider the noise models of the form (2.1) and (2.2) and let $\tilde{\mathbb{Y}}$ be a Hilbert space containing a white noise process W and a measurable estimator $R : \tilde{\mathbb{Y}} \to \mathbb{X}$. Then the worst case risk of R on a set $\mathcal{K} \subset \mathbb{X}$ is defined by*

$$\mathbf{S}_R(\xi,\varepsilon W,\mathcal{K}) := \sup_{\|\xi\| \leq 1} \sup_{f \in \mathcal{K}} \mathbf{E}\left[\left\| R(Tf^\dagger + \xi + \varepsilon W) - f^\dagger \right\|^2 \right]^{1/2}.$$

*The **minimax worst case risk** and **minimax linear worst case risk** are defined by*

$$\mathbf{S}^N(\xi,\varepsilon W,\mathcal{K}) := \inf_R \mathbf{S}_R(\xi,\varepsilon W,\mathcal{K}), \quad \mathbf{S}^L(\xi,\varepsilon W,\mathcal{K}) := \inf_{R,\ linear} \mathbf{S}_R(\xi,\varepsilon W,\mathcal{K}),$$

where the infimum is taken over all estimators R. We say that R is linearly order optimal on $\mathcal{K}$ for the noise model (2.2) if there exist constants $C, \delta_0 > 0$ such that

$$\mathbf{S}(\delta,\varepsilon W,\mathcal{K}) \leq C\mathbf{S}^L(\delta,\varepsilon W,\mathcal{K})$$

for all $\delta \in (0,\delta_0]$ and $\varepsilon > 0$.

Here we note that $\mathbf{S}_R(0,\varepsilon W,\mathcal{K}) = \mathbf{S}_R(\varepsilon W,\mathcal{K})$, $\mathbf{S}_R^L(0,\varepsilon W,\mathcal{K}) = \mathbf{S}_R^L(\varepsilon W,\mathcal{K})$ and $\mathbf{S}_R^N(0,\varepsilon W,\mathcal{K}) = \mathbf{S}_R^L(\varepsilon W,\mathcal{K})$.

First of all, we consider ellipsoids $\mathcal{K}$ of the form by

$$\mathcal{K} = \{f \in \mathbb{X} : \sum_{j=1}^{\infty} \hat{f}_j^2 \leq \rho^2\}. \tag{3.23}$$

with a sequence $\{a_j\}$ and $\lim_{j\to\infty} a_j = \infty$, $a_j > 0$. For this situation, we first need to consider diagonal estimators

$$R_\lambda(f) := \sum_{j=1}^\infty \frac{\lambda_j}{\sigma_j} \langle f, v_j \rangle u_j \tag{3.24}$$

for some sequence $\{\lambda_j\} \in \ell^2(\mathbb{N})$. The risk in the white noise model (2.2) follows from the bias-variance decomposition that

$$\mathbf{S}_{R_\lambda}(\varepsilon W, f) = \mathbf{E}\left[\|R_\lambda T f - f\|^2\right] = \sum_{j=1}^\infty \left[(1-\lambda_j)^2 \langle f, u_j \rangle^2 + \varepsilon^2 \frac{\lambda_j^2}{\sigma_j^2}\right]$$

If we set $\hat{f}_j := a_j \langle f, u_j \rangle$ then we have

$$\mathbf{S}_{R_\lambda}(\varepsilon W, \mathcal{K}) = \rho^2 \sup_{j\in\mathbb{N}} \frac{(1-\lambda_j)^2}{a_j^2} + \varepsilon^2 \frac{\lambda_j^2}{\sigma_j^2}.$$

Theorem 3.9 (Pinsker estimator). *Consider a sequence $(a_j)_{j\in\mathbb{N}}$ such that $a_j > 0$ and $\lim_{j\to\infty} a_j = 0$ and let $\mathcal{K}$ be given by (3.23). Then there exists a unique minimax linear estimator on $\mathcal{K}$. It is given by the operator $R_{\bar{\lambda}}$ defined in (3.24) with the Pinsker weights*

$$\bar{\lambda}_j = \max(1 - \bar{\kappa} a_j, 0)$$

where the constant $\bar{\kappa} > 0$ is the unique solution to the equation

$$\kappa\rho^2 - \varepsilon^2 \sum_{j=0}^\infty \frac{a_j}{\sigma_j} \max(1 - \kappa a_j, 0).$$

The minimax linear risk is given by

$$\mathbf{S}^L(\varepsilon W, \mathcal{K}) = (\bar{\kappa}\rho)^2 + \varepsilon^2 \sum_{j=1}^\infty \frac{\bar{\lambda}_j^2}{\sigma_j^2}.$$

Proof. See Theorem 4.2 [34]. $\qquad\square$

We now compare the linear worst case risk of Pinsker estimator to the spectral cut-off estimator, because the latter one is known to be linearly order optimal. The linear order optimality of spectral regularization methods is given by in the following theorem, which has been proved by Hohage. [1]

Theorem 3.10. *Let W be a white noise process with $\mathbf{Cov}[W] = I$ and $\mathbf{E}[W] = 0$ and $\mathcal{K}_{\varphi,\rho}$ given by (3.5). Assume that the estimator $R_{\alpha_P}^{sc} := q_{\alpha_P}^{sc}(T^*T)T^*$ with $\alpha_P := \varphi^{-1}(2\bar{\kappa})$ and $\bar{\kappa}$ is the unique solution as defined in Theorem 3.9. Then the risk of R_{α_P} satisfies the order optimal risk bound*

$$\mathbf{S}_{R_{\alpha_P}^{sc}}(\varepsilon W, \mathcal{K}_{\varphi,\rho}) \leq 4\mathbf{S}^L(\varepsilon W, \mathcal{K}_{\varphi,\rho}) \tag{3.25}$$

Proof. We set $J := \max\{j : \varphi(\sigma_j^2) \geq 2\bar{\kappa}\}$, the bias and variance term are bounded by

$$\sup_{f\in\mathcal{K}_{\varphi,\rho}} \left\|(I - R_{\alpha_P}^{sc} T)f\right\| \leq \varphi(\alpha_P)\rho, \quad \mathbf{Cov}\left[R_{\alpha_P}^{sc} W\right] = \sum_{j=1}^J \sigma_j^{-2}.$$

---

[1]The theorem 3.10 and its proof are cited from unpublished lecture notes on Chapter 7 of Thorsten Hohage, I would like to thank him to provide the important result to make my thesis more readable.

The Pinsker weights satisfy that $\bar{\lambda}_j = 1 - \frac{\bar{\kappa}}{\varphi(\sigma_j^2)} \geq \frac{1}{2}$ for all $j \leq J$. Therefore,

$$\mathbf{E}\left[\left\|R_{\alpha_P}^{sc}(Tf + \varepsilon W) - f\right\|^2\right] \leq \varphi(\alpha_P)^2 \rho^2 + \varepsilon^2 \sum_{j=1}^{J} \sigma_j^{-2}$$

$$\leq 4\left(\bar{\kappa}^2 \rho^2 + \varepsilon^2 \sum_{j=1}^{\infty} \frac{\bar{\lambda}_j^2}{\sigma_j^2}\right) = 4\mathbf{S}^L(\varepsilon W, \mathcal{K}_{\varphi,\rho})$$

for all $f \in \mathcal{K}_{\varphi,\rho}$. $\square$

Theorem 3.10 shows that the estimator R on the set $\mathcal{K}_{\varphi,\rho}$ linearly order optimal up to some constant. It has been known that spectral cut-off estimators are order optimal in a minimax sense under some conditions, see [64]. From Theorem 6 in Bissantz et al [11] we further know that general spectral regularization methods can achieve the same rates of convergence as spectral cut-off if the optimal value of α is chosen properly for every noise level ε. Therefore, the convergence rates (3.19) and (3.22) are linearly order optimal.

It is also interesting to highlight that a comparison of the linear minimax risk $\mathcal{S}^L$ with the nonlinear risk are studied by Pinsker [73], where he showed under the additional assumptions that W is Gaussian and

$$\sup_{J \in \mathbb{N}} \frac{\sum_{j=1}^{J} \sigma_j^2}{\sup_{j \leq J} \sigma_j^2} < \infty$$

then $\mathbf{S}^L(\varepsilon W, \mathcal{K}) \sim \mathbf{S}^N(\varepsilon W, \mathcal{K})$ as noise level tends to zero.

3.2   Tikhonov regularization

In this section, we deal with the results on convergence rates of Tikhonov regularization for the solution of linear ill-posed inverse problem

$$Tf^\dagger = g^\dagger. \tag{3.26}$$

One of the most common approaches to compute a stable approaximation of $f^\dagger$ given g^{obs} is Tikhonov regularization of the form

$$\widehat{f}_\alpha \in \underset{f \in \mathbb{X}}{\arg\min} \left[\frac{1}{2}\|Tf - g^{\text{obs}}\|_{\mathbb{Y}}^2 + \frac{\alpha}{2}\|f - f_0\|_{\mathbb{X}}^2\right], \tag{3.27}$$

where $\alpha > 0$ is called regularization parameter, which is determining the weight of penalty terms. If initial guess is unknown, then we take $f_0 = 0$. By choosing the initial guess in a proper way, a-*priori* information about the unknown solution $f^\dagger$ may be incorporated into the regularization procedure. The form (3.27) has been considered by Russian mathematician Tikhonov [87], which is suitable for classical setting with the squared norm of $\mathcal{S}(g, g^{\text{obs}}) = \|g - g^{\text{obs}}\|_{\mathbb{Y}}^2$ and $\mathcal{R}(f) = \|f - f_0\|_{\mathbb{X}}^2$. Thus, the minimizer $\widehat{f}_\alpha$ in (3.27) can be computed by explicitly for a bounded linear operator T, i.e.,

$$\widehat{f}_\alpha = (T^*T + \alpha I)^{-1}(T^* g^{\text{obs}} + \alpha f_0), \tag{3.28}$$

see Theorem 5.1 [28]. Our aim in this section is to study convergence analysis for Tikhonov regularization in a stochastic setting. Since white noise on a Hilbert space

$\mathbb{Y}$ does not belong to $\mathbb{Y}$ with probability 1, by expanding the data fidelity term in (3.27) and subtracting the term $\frac{1}{2}\|g^{\text{obs}}\|_{\mathbb{Y}}^2$ as well as take the initial guess $f_0 = 0$, the minimization problem (3.27) can be reduced to the following form

$$\widehat{f}_\alpha \in \operatorname*{argmin}_{f \in \mathbb{X}} \left[\frac{1}{2}\|Tf\|_{\mathbb{Y}}^2 - \langle Tf, g^{\text{obs}} \rangle + \frac{\alpha}{2}\|f\|_{\mathbb{X}}^2 \right]. \tag{3.29}$$

Convergence rates of Tikhonov regularization for linear and nonlinear inverse problems under variational source conditions have been studied in many publications (see e.g., [31, 36, 42, 78, 81, 93]), where they studied the convergence analysis under the deterministic and stochastic framework. In particular, Engl et al [28] derived optimal convergence rates for Tikhonov regularization with linear operator under Hölder-type source conditions and deterministic assumptions. Burger et al [17] obtained nearly optimal convergence rates for variational regularization with Gaussian white noise. In this setion we mainly consider the optimal convergence rates for Gaussian white noise under Hölder-type source conditions. Futhermore, the data fidelity term $\mathcal{S}$ and penalty term $\mathcal{R}$ can also be generalized to nonquadratic cases including norm powers, maximum entropy, and ℓ^1- type sparsity, we will dicuss generalized Tikhonov regularization with nonquadratic data fidelity terms $\mathcal{S}$ for Poisson and empirical process data in Chapter 5.

3.2.1 Error bounds in expectation

Before we present a convergence analysis for the regularization method (3.29), we would like to recall a new concept called **effective noise level**, which was introduced first time by Werner and Hohage [95]. Since the observed data g^{obs} are only got access through the data fidelity term when the variational regularization methods are considered, it is a main procedure to quantify the influence of the data g^{obs} on the functional $\mathcal{S}$.

To minimize $\mathcal{S}(g^{\text{obs}}, g)$, we are interested in the minimum value of $\mathcal{S}(g^{\text{obs}}, g)$. Due to the free additive constant in $\mathcal{S}(g^{\text{obs}}, g)$, the value of minimum has no significance. Thus, we can subtract $\mathcal{S}(g^{\text{obs}}, g^\dagger)$. Moreover, it is also important to ensure that the distance between $\mathcal{S}(g^{\text{obs}}, g)$ and $\mathcal{S}(g^{\text{obs}}, g^\dagger)$ is arbitrary large if g is far from $g^\dagger$. Therefore, we introduce a non-negative data fidelity functional $\mathcal{T}_{g^\dagger} : \mathbb{Y} \to [0, \infty)$ for exact data $g^\dagger$ and $T(g^\dagger, g) = 0$ if and only if $g^\dagger = g$.

Definition 3.11 (Effective noise level). *Let $g^\dagger \in \mathbb{Y}$ be denoted the exact data, g^{obs} the observed data. Let $C_{err} \geq 1$ be a constant. Then the effective noise level at any point $g \in \mathbb{Y}$ is defined by*

$$\mathbf{err}(g) := \frac{1}{C_{err}}\mathcal{T}(g, g^\dagger) - \mathcal{S}(g^{obs}, g) + \mathcal{S}(g^\dagger, g^{obs})$$

and a two-sided effective noise level at $g \in \mathbb{Y}$ defined by

$$\overline{\mathbf{err}}_{\pm}(g) := \max\left\{ \mathbf{err}(g), \frac{1}{C_{err}}\mathcal{T}(g, g^\dagger) - \mathcal{S}(g^{obs}, g) + \mathcal{S}(g^\dagger, g^{obs}) \right\}.$$

The two-sided effective noise level will be expected when we discuss iteratively regularized Newton-type methods in Chapter 5. We will consider that $\mathbf{err}(g)$ is not only finite but also sufficiently small for $g \in \mathbb{Y}$. For the classical deterministic models, we assume that $\|g^{\text{obs}} - g^\dagger\|_{\mathbb{Y}} \leq \delta$ and

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

for $r \geq 1$, then we obtain from the inequalities

$$(a+b)^r \leq 2^{r-1}(a^r + b^r) \quad \text{and} \quad |a-b|^r + b^r \geq 2^{1-r}a^r$$

that $\mathbf{err}(g) = 2\|g^{\text{obs}} - g^\dagger\|_{\mathbb{Y}}$ hold true with constant $C_{err} = 2^{r-1}$. For Gaussian white noise we set $C_{err} = 1$ in Definition 3.11 such that the data noise error can be computed by

$$
\begin{aligned}
\mathbf{err}(g) :&= \mathcal{S}(g^{\text{obs}}, g^\dagger) - \mathcal{S}(g^{\text{obs}}, g) + \mathcal{T}(g^\dagger, g) \\
&= \frac{1}{2}\|g^\dagger\|^2 - \langle g^{\text{obs}}, g^\dagger \rangle - \frac{1}{2}\|g\|^2 + \langle g^{\text{obs}}, g \rangle + \frac{1}{2}\|g^\dagger - g\|_{\mathbb{Y}}^2 \\
&= \langle \varepsilon W, g - g^\dagger \rangle.
\end{aligned}
\tag{3.30}
$$

Now we need to bound the noise error term at the point $g = T\widehat{f}_\alpha$ for $g \in \mathbb{Y}$.

Theorem 3.12. *Let $T : \mathbb{X} \to \mathbb{Y}$ be a Hilbert-Schmidt operator in $\mathbb{Y}$. Suppose that Assumptions 3.1 and 3.4 hold true. We define a general spectral estimator $\widehat{f}_\alpha$ by (3.29). Then we obtain*

$$\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right] \leq \widetilde{C}\varepsilon^2 \widetilde{\Gamma}(\alpha)$$

for all $\alpha > 0$ with some constant $\widetilde{C} > 0$, which is given by

$$\widetilde{C} = \left(2h + C_\varphi + 2hC_\varphi\right)C_{q,1}^2 + (1+h)C_{q,2}^2\right).$$

Proof. Set $\widehat{f}_\alpha = R_\alpha(Tf^\dagger + \varepsilon W)$, then for any $T\widehat{f}_\alpha \in \mathbb{Y}$, we have

$$
\begin{aligned}
g = T\widehat{f}_\alpha &= TR_\alpha(g^\dagger + \varepsilon W) = T(T^*T + \alpha I)^{-1}T^*(g^\dagger + \varepsilon W) \\
&= T(T^*T + \alpha I)^{-1}T^*g^\dagger + \varepsilon T(T^*T + \alpha I)^{-1}T^*W.
\end{aligned}
\tag{3.31}
$$

By inserting (3.31) into (3.30), we have

$$
\begin{aligned}
\mathbf{err}(T\widehat{f}_\alpha) :&= \langle \varepsilon W, g - g^\dagger \rangle_{\mathbb{Y}} = \varepsilon \langle W, T\widehat{f}_\alpha - Tf^\dagger \rangle_{\mathbb{Y}} \\
&= \varepsilon \langle W, T(T^*T + \alpha I)^{-1}T^*Tf^\dagger - Tf^\dagger \rangle_{\mathbb{Y}} \\
&\quad + \varepsilon^2 \langle W, (T^*T + \alpha I)^{-1}T^*TW \rangle_{\mathbb{Y}}.
\end{aligned}
\tag{3.32}
$$

Since the first term on the right hand side of (3.32) is deterministic, i.e., $T(T^*T + \alpha I)^{-1}T^*Tf^\dagger - Tf^\dagger \in \mathbb{Y}$ by taking an expectation on both side of (3.32) and from the property of white noise process, we obtain

$$\mathbf{E}[\langle W, T(T^*T + \alpha I)^{-1}T^*Tf^\dagger - Tf^\dagger \rangle_{\mathbb{Y}}] = 0.$$

Now it remains to bound on the second term of (3.32). By Assumption 3.1 we have

$$
\begin{aligned}
\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right] &= \varepsilon^2 \mathbf{E}\langle W, (T^*T + \alpha I)^{-1}T^*TW \rangle_{\mathbb{Y}} = \varepsilon^2 \text{Tr}(q_\alpha(T^*T)T^*T) \\
&\leq \varepsilon^2 \sum_{j=1}^{\infty} q_\alpha(\sigma_j^2)\sigma_j^2 = -\varepsilon^2 \int_0^\infty q_\alpha(\theta)\theta \mathrm{d}\Gamma(\theta) \\
&\leq -\varepsilon^2 \frac{C_{q,2}}{\alpha}\int_0^\alpha \theta \mathrm{d}\Gamma(\theta) - C_{q,1}\varepsilon^2 \int_\alpha^\infty \mathrm{d}\Gamma(\theta),
\end{aligned}
\tag{3.33}
$$

where $q_\alpha(T^*T) := (\alpha I + T^*T)^{-1}$. If Assumptions 3.1 and 3.4 hold true, then using Lemma 12 and Lemma 13 in Bissantz et al [11], we obtain

$$\text{Tr}(T^* q_\alpha(T^*T)T) \leq -\frac{C_{q,2}}{\alpha} \int_0^\alpha \theta d\Gamma(\theta) - C_{q,1} \int_\alpha^\infty d\Gamma(\theta)$$
$$\leq \left((2h + C_\Gamma + 2hC_\Gamma)C_{q,1}^2 + (1+h)C_{q,2}^2\right)\widetilde{\Gamma}(\alpha).$$

Thus, by plugging this inequality into (3.33), we derive

$$\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right] \leq \widetilde{C}\varepsilon^2\widetilde{\Gamma}(\alpha),$$

where $\widetilde{C} = \left((2h + C_\Gamma + 2hC_\Gamma)C_{q,1}^2 + (1+h)C_{q,2}^2\right)$, this complete the proof of theorem. $\qquad\square$

Corollary 3.13. *Let* $T : \mathbb{X} \to \mathbb{Y}$ *be a Hilbert-Schmidt operator in* $\mathbb{Y}$. *Suppose that Assumptions 3.1 and 3.4 hold true and the sigular values* σ_j *of T are given by* $\sigma_j(T) = j^{-p}$ *for* $p > 1/2$. *The spectral estimator* $\widehat{f}_\alpha$ *is defined by (3.29). Then we obtain*

$$\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right] \leq \widetilde{C}\varepsilon^2 \alpha^{-\frac{1}{2p}},$$

for all $\alpha > 0$, *where* $\widetilde{C} = \left(\frac{C_{q,2}(2-C_\alpha)}{2p-1} - C_{q,1}C_\alpha\right)$.

Proof. This is a direct consequence of Theorem 3.12 with the choice $\widetilde{\Gamma}(\alpha) = \alpha^{-\frac{1}{2p}}$ such that the proof is omitted. $\qquad\square$

Analogously, we can also bound the noise error in expectation for exponentially ill-posed inverse problems such as backwards heat equation, where the logarithmic source conditions $\varphi_p = (-\ln\lambda)^{-p}$ for $p > 0$ are more appropriate.

Corollary 3.14. *Let* $T : \mathbb{X} \to \mathbb{Y}$ *be a Hilbert-Schmidt operator in* $\mathbb{Y}$. *Suppose that Assumptions 3.1 and 3.4 hold true and the sigular values* σ_j *of T are exponentially decaying, which are given by* $\sigma_j(T) = \exp(-2t\lambda_j)$ *for* $p > 0$. *The spectral estimator* $\widehat{f}_\alpha$ *is defined by (3.29). Then we obtain*

$$\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right] \leq \widetilde{C}c\varepsilon^2 \left(-\frac{1}{2t}\ln\alpha\right)^{d/2},$$

for all $\alpha > 0$, *where* $\widetilde{C} = \left(\frac{C_{q,2}(2-C_\alpha)}{2p-1} - C_{q,1}C_\alpha\right)$.

Proof. This is a direct consequence of Theorem 3.12 with the choice

$$\widetilde{\Gamma}(\alpha) = c\left(-\frac{1}{2t}\ln\alpha\right)^{d/2}$$

such that the proof is omitted. $\qquad\square$

3.2.2 Variational source conditions

In this section we will recall the regularization theory for variational methods. To estimate the smoothness of the unknown solution, we always use some sort of source conditions, which can be written in the form of a variational inequality introduced by Hofmann et al [42]. Since the derivation of convergence rates for the

inverse problem requires a-priori knowledge about the unknown solution $f^\dagger$, we usually measure it by Bregman distance. However, the distance measure with a norm is too restrictive to analyze convergence rates in the general cases (1.3). Now we will give a precise definition of subgradient for a convex function.

Definition 3.15 (**Subgradient**)**.** *Let $\mathbb{X}$ and $\mathbb{X}^*$ be denote locally convex vector space and its dual space. $f^* \in \mathbb{X}^*$ is called a subgradient of a convex functional $\mathcal{R} : \mathbb{X} \to \bar{\mathbb{R}}$ at $f \in \mathbb{X}$ if $\mathcal{R}(f)$ is finite and*

$$\mathcal{R}(\widetilde{f}) \geq \mathcal{R}(f) + \langle f^*, \widetilde{f} - f \rangle \quad \text{for all } \widetilde{f} \in \mathbb{X}.$$

The set of all subgradients of $\mathcal{R}$ at f is called the subdifferential $\partial \mathcal{R}(f)$.

Definition 3.16 (**Bregman distance**)**.** *Let $\mathcal{R} : \mathbb{X} \to (-\infty, \infty]$ is a proper, convex, lower-semicontinuous functional. For a subgradient $f^* \in \partial \mathcal{R}(f^\dagger)$, the **Bregman distance** of $\mathcal{R}$ is given by*

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

for $f \in \mathbb{X}$, where $\langle \cdot, \cdot \rangle$ denotes the duality product of $\mathbb{X}^*$ and $\mathbb{X}$.

Recently Bregman distances became a widely used tool in inverse problems since they are related to the log-likelihood of the density of exponential families [4, 75]. There exists a good connection between exponential families and Bregman distances. For example, the Poisson distributions correspond to generalized Kullback-Leibler divergence, exponential distributions correspond to the Itakura-Saito divergence etc. The Bergman distance was first time introduced in the context of inverse problems by Eggermont [26], he analyzed such problems with the maximum-entropy regularization and later Burger and Osher [16] studied convergence rates for convex variational regularization by using a general penalization term. It has been known that the Bregman distance is nonnegative due to convexity and $\mathcal{D}_{\mathcal{R}}^{f^*}(f^\dagger, f^\dagger) = 0$ if and only if $f = f^\dagger$. In fact, the Bregman distance is not a real distance since it is neither symmetric nor does it satisfy a triangle inequality, which depends on the penalty term $\mathcal{R}$ and the choice of f^*. If $\mathcal{R}(f) = \frac{1}{2}\|f - f_0\|_{\mathbb{X}}^2$ in Hilbert spaces, then we have $\mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger) = \frac{1}{2}\|f - f^\dagger\|_{\mathbb{X}}^2$.

Moreover, it is worth to emphasize that the relationship between the Bregman distance and a metric space, which is a particularly important result to derive convergence rates in terms of the norm.

Lemma 3.17. *Let $\mathbb{X}$ be a q-convex Banach space and the penalty function $\mathbb{R}(f) = \frac{1}{q}\|f\|_{\mathbb{X}}^q$. Then there exists some constant $C_{bd} > 0$ such that*

$$\|f - f^\dagger\|_{\mathbb{X}}^q \leq C_{bd}\mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger) \tag{3.34}$$

for all $f \in \mathcal{B} \subset \mathbb{X}$.

Proof. See Xu and Roach [97] or Sprung [83] for a simplified proof. □

This is an essential result since then convergence and convergence rates with respect to the Bregman distance imply convergence and convergence rates with respect to norm powers. The inequality (3.34) is also valid for different types of penalty terms $\mathcal{R}$ such as entropy functional $\mathcal{R}_{ME}$ and total variation functional $\mathcal{R}_{TV}$. For example, the Bregman distance $\mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger)$ is exactly the Kullback-Leibler divergence for the entropy functional $\mathcal{R}$.
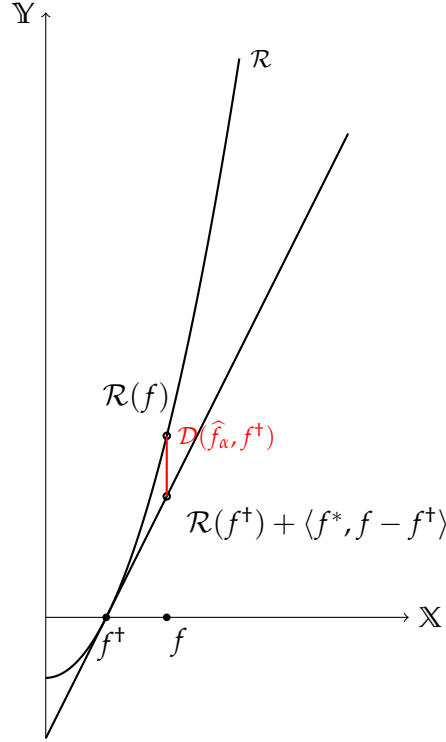
FIGURE 3.1: The Bregman distance of $\mathcal{R}(x) = x^2 - 1$ between f and $f^\dagger$

Under mild assumptions it can be shown that the Bregman distance $\mathcal{D}_\mathcal{R}^{f^*}(\widehat{f}_\alpha, f^\dagger)$ of variational estimators $\widehat{f}_\alpha$ converges to 0 in probability as the noise level tends to zero, however convergence rates can be arbitrarily slow in the context of ill-posed inverse problems, see e.g. [53]. To derive a specific convergence rate we need additional assumptions on the true solution $f^\dagger$. Such smoothness assumptions were introduced by Hofmann et al [42] and it became an increasingly popular tool in inverse problems in Banach space setting [15, 16, 32, 31, 36, 43, 93]. Now we formulate our assumption on the smoothness of the solution $f^\dagger$ for general cases.

Assumption 3.18 (Varitional source condition). *Let $\mathcal{R} : \mathbb{X} \to (-\infty, \infty]$ be a proper, convex and semicontinuous functional and there exist $f^* \in \partial \mathcal{R}(f^\dagger)$. Assume that the variational inequality*

$$\beta \mathcal{D}_\mathcal{R}^{f^*}(f, f^\dagger) \leq \mathcal{R}(f) - \mathcal{R}(f^\dagger) + \Phi\left(\mathcal{T}_\zeta(g^\dagger, F(f))\right) \tag{3.35}$$

for all $f \in \mathbb{X}$ and a constant $\beta > 0$ holds true, where the index function $\Phi : [0, \infty) \to [0, \infty)$ is concave monotonically increasing and $\Phi(0) = 0$.

A variational source condition with the quadratic penalty functional $\mathcal{R}$ and the quadractic exact data fidelity functional $\mathcal{T}$ in Hilbert spaces, then the variational inequality (3.35) can be derived from

$$f^\dagger - f_0 = \Psi(F'[f^\dagger]^* F'[f^\dagger])\omega \tag{3.36}$$

for some $\omega \in \mathbb{X}$ with the tangential cone condition

$$\|F(f) - F(f^\dagger) - F'[f^\dagger](f - f^\dagger)\|_\mathbb{Y} \leq C\|F(f) - F(f^\dagger)\|_\mathbb{Y} \tag{3.37}$$

for all $f \in \mathbb{X}$ and a constant $C > 0$, where $\Psi : [0, \infty) \to [0, \infty)$ is also index function and Ψ^2 is concave, $F : \mathcal{D}(F) \subset \mathbb{X} \to \mathbb{Y}$ is Fréchet differentiable in an open domain $\mathcal{D}(F)$. In fact (3.36) and (3.37) implies the following variational inequality

$$\frac{\beta}{2} \|f - f^\dagger\|_{\mathbb{X}}^2 \leq \frac{1}{2} \|f\|_{\mathbb{X}}^2 - \frac{1}{2} \|f^\dagger\|_{\mathbb{X}}^2 + \Phi\left(\|F(f) - g^\dagger\|_{\mathbb{Y}}^2\right) \tag{3.38}$$

for a constant $\beta > 0$ and $f \in \mathcal{D}(F)$, see Flemming [32, 31]. Futhermore, Flemming also shows that a spectral source condition implies a variational source condition in [31]. Obviously, Assumption 3.18 holds true for some index function $\Phi = \Phi_\Psi$ for linear operator $F = T$ and $\mathcal{T}(g^\dagger, g) = \|g - g^\dagger\|_{\mathbb{Y}}^2$, see [32, 33]. For instance,

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

and

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

For more details of the proof, we refer to [32]. In Hilbert space setting, the variational source conditions are weaker than spectral source conditions. Weidling and Hohage [50] also point out the many advantages of the variational source conditions compare to the spectral source conditions.

The variational inequality can be used as the source condition to the inverse problems with Gaussian white noise in Hilbert spaces, see Chapter 5. Moreover, it is easy to see that (3.38) implies

$$\frac{\beta}{2} \|f - f^\dagger\|_{\mathbb{X}}^2 \leq \Phi(\|F(f) - F(f^\dagger)\|_{\mathbb{X}}^2)$$

for all $f, f^\dagger \in \mathcal{B} \subset \mathbb{X}$. Nevertheless, Frederic and Hohage [49] show that the stability estimates for an inverse scattering problems can be sharpened to variational source conditions.

Apart from this, there exist other types of varional inequalities used as source conditions. In particular, to overcome the computational difficulties of functional calculus for the range type conditions (3.36), Hofmann et al [42] introduced a additive form of variational inequality. There exists a subgradient $f^* \in \partial \mathcal{R}(f^\dagger)$, a parameter $\beta \in [0, 1)$ and an index funtion Φ such that

$$\langle f^*, f^\dagger - f \rangle \leq \beta \mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger) + \Phi_{\text{add}}(\mathcal{S}(F(f), g^\dagger)) \tag{3.39}$$

for all $f \in \mathcal{B}$. This is equivalent (3.38) with $\Phi = \Phi_{\text{add}}$. Scherzer et al [78] derived convergence rates for Tikhonov-type regularization by using (3.39) as the source conditions. Hofmann and Yamamoto [43] proved the same role of the source conditions (3.36) and (3.39) with index function $\Phi_{\text{add}} = \Psi_{\frac{1}{2}}$ in Hilbert space case and show that the conditions (3.39) is limited to the case that Φ_{add}^2 is a concave function. Moreover, Pöschl [75] derived convergence rates for Tikhonov-type regularization where the data fidelity functionals given by non-qudratic norm powers. Flemming [32, 31] even proved convergence rates for Tikhonov-type regularization with general type of functionals $\mathcal{S}$ and $\mathcal{R}$. Later Werner [94] used (3.35) to deal with convergence rates for Poisson data. No Fréchet derivative is required nonlinear F in (3.38). In addition,

Kaltenbacher and Hofmann [56] proposed a multiplicative form of a variational inequality as a source condition, which is given by

$$\langle f^*, f^\dagger - f \rangle \leq \beta \mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger)^{\frac{1}{2}} \Phi_{\text{mult}} \left(\frac{\mathcal{S}(F(f), g^\dagger)}{\mathcal{D}_{\mathcal{R}}^{f^*}(f, f^\dagger)} \right) \tag{3.40}$$

for all $f \in \mathcal{B}$. Moreover, we assume Φ_{mult} is such that

$$r \mapsto \frac{\Phi_{\text{mult}}(r)}{\sqrt{r}} \tag{3.41}$$

is monotonically increasing, where the condition (3.41) restricts the the index functions Φ_{mult} is concave as the function Φ_{add}^2 in (3.39). For example, $\Phi_{\text{mult}}(r) = r^\nu$ with $\nu \leq \frac{1}{2}$ for Hölder-type variational inequalities. Werner in Lemma 3.20 [94] show that (3.40) is valid for the quadratic form of data fidelity and penalty functionals in Hilbert spaces. Moreover, Hofmann and Yamamoto [43] prove that $\Phi_{\text{add}}(r) = r^\nu$ for additive variational inequality (3.39), which is also limited to $\nu \leq \frac{1}{2}$ if $f^* \neq 0$.

Definition 3.19 (Conjugate function). *Let $\Phi : (-\infty, \infty) \to (-\infty, \infty]$ be a function defined on $\mathbb{R}$. Then the Fenchel conjugate Φ^* is defined by*

$$\Phi^*(\gamma) = \sup_{\omega \geq 0} (\gamma\omega - \Phi(\omega)), \quad \gamma \in \mathbb{R}. \tag{3.42}$$

The conjugate function Φ^* is convex in general as supremum over the affine linear function $\gamma \mapsto \gamma\omega - \Phi(\omega)$. As given by Grasmair [36], the bias (approximation error) will be bounded by the function $\Psi : (0, \infty) \to [0, \infty]$, i.e.,

$$\Psi(\gamma) = (-\Phi)^*(-\frac{1}{\gamma}).$$

For details on the Fenchel duality, we refer to [27]. Deterministic convergence rates results for ill-posed inverse problems have been obtained with different noise models, (see [31, 32, 52, 53, 95]). By observing the proof of Theorem 5.1 in [95], we know that the index function $\Psi(\gamma)$ is monotonically increasing. For all $\gamma > 0$ and $C > 0$, it holds that

$$\Psi(C\gamma) = (-\Phi)^* \left(-\frac{1}{C\gamma} \right) \leq \max\{1, C\}(-\Phi)^* \left(-\frac{1}{\gamma} \right) = C\Psi(\gamma).$$

Lemma 3.20. *Assume that the index function Φ satisfy $\Phi(\alpha) = \beta\alpha^{\frac{2\nu}{2\nu+1}}$ for some constant $\beta > 0$ and $\nu \in (0, 1/2)$. Then we obtain the Fenchel conjugation of Φ by*

$$(-\Phi)^*(\gamma) = \begin{cases} \infty, & \text{if } \gamma \geq 0, \\ C_{\nu\beta}\gamma^{-2\nu}, & \text{if } \gamma < 0 \end{cases} \tag{3.43}$$

with some constant $C_{\nu\beta} = \left[-\frac{2\nu+1}{2\nu\beta} + \beta \left(\frac{2\nu+1}{2\nu\beta} \right)^{\frac{2\nu}{2\nu+1}} \right] > 0$.

Proof. Since Φ is concave function and continuous, $-\Phi$ is convex such that

$$0 = \Phi^*(\tau^*)' = \gamma + \frac{2\nu\beta}{2\nu+1}(\tau^*)^{-\frac{1}{2\nu+1}} \Rightarrow \tau^* = -\frac{2\nu+1}{2\nu\beta}\gamma^{-2\nu-1}. \tag{3.44}$$

By inserting (3.44) into the conjugate function (3.42), we derive

$$
\begin{aligned}
(-\Phi)^*(\gamma) &= \left[\gamma \tau^* + \Phi(\tau^*) \right] \\
&= \left[-\frac{2\nu+1}{2\nu\beta} \gamma^{-2\nu-1} \cdot \gamma + \beta \left(-\frac{2\nu+1}{2\nu\beta} \gamma^{-2\nu-1} \right)^{\frac{2\nu}{2\nu+1}} \right] \\
&= \left[-\frac{2\nu+1}{2\nu\beta} \gamma^{-2\nu} + \beta \left(\frac{2\nu+1}{2\nu\beta} \right)^{\frac{2\nu}{2\nu+1}} \gamma^{-2\nu} \right] \\
&= \left[-\frac{2\nu+1}{2\nu\beta} + \beta \left(\frac{2\nu+1}{2\nu\beta} \right)^{\frac{2\nu}{2\nu+1}} \right] \gamma^{-2\nu}
\end{aligned}
\tag{3.45}
$$

for $\gamma < 0$. Thus, the conjugate function of Φ is given by

$$
(-\Phi)^*(\gamma) = \begin{cases} \infty, & \text{if } \gamma \geq 0, \\ \left[-\frac{2\nu+1}{2\nu\beta} + \beta \left(\frac{2\nu+1}{2\nu\beta} \right)^{\frac{2\nu}{2\nu+1}} \right] \gamma^{-2\nu}, & \text{if } \gamma < 0. \end{cases}
\tag{3.46}
$$

Taking $C_{\nu\beta} = \left[-\frac{2\nu+1}{2\nu\beta} + \beta \left(\frac{2\nu+1}{2\nu\beta} \right)^{\frac{2\nu}{2\nu+1}} \right] > 0$ we obtain (3.43). $\qquad\square$

Similarly, the Fenchel conjugate of the convex function $\Phi(\lambda) = \beta(-\ln(\lambda))^{-2p}$ for $p > 0$ has been derived by Flemming [31], which is given by

$$
(-\Phi)^*(\gamma) = \widehat{\beta}(-\ln \lambda)^{-2p}(1 + o(1)), \quad \text{for } p > 0.
\tag{3.47}
$$

3.2.3   Advantages of variational regularization method

In this section we will list some important advantages of variational regularization over the spectral regularization methods. In the last few decades researchers widely studied the regularization methods based on the spectral theory in Hilbert spaces under deterministic and stochastic error assumptions, we refer to [28] for spectral regularization under deterministic setting and other authors [10, 11, 30, 48]. We will list out in the following some important advantages of the convergence analysis in variational regularization compared to the analysis of the estimator (3.1) based on the spectral theory.

(a) Variational regularization methods can be applied to inverse problems in Banach space settings. For example Flemmming [31], Pöschl [75] and Werner and Hohage [51, 53, 95] widely studied convergence rates for generalized Tikhonov regularization of inverse problems in Banach space setting. However, general spectral regularization methods are even not well defined as an algorithm in Banach spaces. Therefore, no convergence analysis is applicable for the regularization methods based on spectral theory.

(b) Variational regularization methods allow to use more general penalty functionals. In the last decades researchers tried to replace the penalty functional $\|f - f_0\|_{\mathbb{X}}^2$ by some general convex functional $\mathcal{R} : \mathbb{X} \to (-\infty, \infty]$. The spectral regularization methods can only use for the quadratic form of penalty functional in Hilbert spaces, In variational regularization, the choice of $\mathcal{R}$ include

entropy like functionals of the form

$$\mathcal{R}_{ME}(f) = \int_{\mathbb{M}} f(x) \ln(f(x)) \mathrm{d}x, \tag{3.48}$$

where $u \in \mathbb{X} = L^1(\mathbb{M})$, $\mathbb{M} \subset \mathbb{R}^d$ is some priori information (see Sec. 5.3 in [28]).

Another type of penalty functional is given by the form of total variation semi-norm

$$\mathcal{R}_{TV}(f) = \sup\left\{ \int_{\mathbb{M}} f \mathrm{div} v \mathrm{d}x \Big| v \in C_0^1(\mathbb{M})^d, \|v\|_\infty \le 1 \right\}.$$

If f is an element of Sobolev space $W^{1,1}(\mathbb{M})$, then $\mathcal{R}_{TV}(f) = \int_{\mathbb{M}} |\nabla f| \mathrm{d}x$ (see e.g. [1, 78]). Furthermore, one of the recent developed regularization with sparsity promoting functionals given by a weighted ℓ^1-norm

$$\mathcal{R}_{\ell^1}(f) = \sum_{i \in I} \omega_i |\langle f, \phi_i \rangle|, \quad f \in \mathbb{X}, \tag{3.49}$$

where $\{\phi\}_{i \in I}$ is some orthonormal basis of the Hilbert space $\mathbb{X}$ and $|\langle f, \phi_i \rangle|$ are coefficients of f. Minimization problems with quadratic data fidelity functional $\mathcal{S}$ and sparsity form of penalty functional $\mathcal{R} = \mathcal{R}_{\ell^1}$ has been considered by Daubechies et al [22] and Zarzer [98]. Furthermore, more research work for the regularization has been explored in a number of papers (see e.g., [18, 47, 78].

(c) Variational regularization method is even applicable for inverse problems with general data fidelity functionals. For example, the most general case (1.3) has been explored by Pöschl [75] where the data fidelity functional $\mathcal{S}(g^{\mathrm{obs}}, g^\dagger)$ is assumed to satisfy a triangle inequality under the variational inequality (3.39). Furthermore, inverse problems with Poisson and empirical process data have been studied by Hohage and Werner [51, 53, 95] as well as Dunker and Hohage [24], respectively. For the cases of inverse problems with Poisson and empirical process data, $\mathcal{S}(g^{\mathrm{obs}}, g)$ is essentially given by the Kullback- Leibler divergence. However, this can be treated by generalizing the noise model to the effective noise level given by Definition 3.11. Flemming [32] proved the optimal rates of convergence for Hölder-type and logarithmic source conditions in the quadratic Hilbert space setting, where he considered the problem with generalized data fidelity functionals.

(d) No restrictive assumption on the operator F and its derivative F' is needed in variational regularization methods. This cannot be verified for many important applications in practice.

3.3 Convergence rates result for additive random noise

3.3.1 Deterministic convergence rates result

In this section, we will first recall a general convergence rates result for variational regularizations under a deterministic noise assumption. Werner and Hohage [95] proved the following deterministic convergence results in a Banach space setting in terms of Bregman distance.

Theorem 3.21 (Deterministic convergence rates). *Suppose that Assumption 3.18 hold trues with the form of (3.38) and the Tikhonov functional (1.3) has a global minimizer $\widehat{f}_\alpha$. Then we obtain*

$$\mathcal{D}_\mathcal{R}^{f^*}(\widehat{f}_\alpha, f^\dagger) \leq \frac{2\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + 2(-\Phi)^* \left(-\frac{1}{C_{err}\alpha}\right) \tag{3.50}$$

for all $\alpha > 0$. Moreover, if $\overline{\mathbf{err}}(F(\widehat{f}_\alpha)) =: \sup_{g \in \mathbb{Y}} \mathbf{err}(F(f))$ is finite, then the infimum of the right hand side of (3.50) is attained at $\alpha = \alpha^*$ if and only if α is priori such that

$$-\frac{1}{\alpha^*} \in \partial(-\Phi)(\overline{\mathbf{err}}). \tag{3.51}$$

Then such a choice of α in (3.51), we obtain the following convergence rate

$$\mathcal{D}_\mathcal{R}^{f^*}(\widehat{f}_\alpha, f^\dagger) = O(\Phi(\overline{\mathbf{err}})).$$

Proof. See Theorem 3.3 in [95]. □

Theorem 3.21 is general deterministic convergence result to derive optimal convergence rates for variational regularization, but we only need quadratic distance norm in Hilbert spaces rather than Bregman distance due to the quadratic regularized term $\mathcal{R}$. Now we would like to state the optimal convergence results for Hölder-type and logarithmic source conditions in a deterministic setting in terms of the noise level δ, respectively.

Theorem 3.22 (Convergence rates for Hölder-type source conditions). *Let $\mathbb{X}$ and $\mathbb{Y}$ be Hilbert spaces and T be a bounded linear operator between $\mathbb{X}$ and $\mathbb{Y}$. Assume that the data fidelity functional $\mathcal{S}(g, g^{obs})$ and penalty functional $\mathcal{R}(f)$ are given by quadratic norms, i.e.,*

$$\mathcal{S}(g, g^{obs}) = \frac{1}{2}\|g - g^{obs}\|_\mathbb{Y}^2 \quad and \quad \mathcal{R}(f) = \frac{1}{2}\|f - f_0\|_\mathbb{X}^2.$$

If the range condition (3.36) holds true with $\Psi(\lambda) = \lambda^\nu$ for $\nu \in (0, \frac{1}{2}]$ and a upper bound $\|g^\dagger - g^{obs}\|_\mathbb{Y} \leq \delta$ where the parameter $\alpha > 0$ is chosen as $\alpha = \delta^{\frac{2}{2\nu+1}}$, then we obtain the following convergence rates

$$\left\|\widehat{f}_{\alpha^*} - f^\dagger\right\|_\mathbb{X} = O\left(\delta^{\frac{2\nu}{2\nu+1}}\right) \tag{3.52}$$

for all $\delta \to 0$.

Proof. See Corollary 3.33 in [94]. □

Theorem 3.23 (Convergence rates for logarithmic source conditions). *Let $\mathbb{X}$ and $\mathbb{Y}$ be Hilbert spaces and T be a bounded linear operator between $\mathbb{X}$ and $\mathbb{Y}$. Assume that the data fidelity functional $\mathcal{S}(g, g^{obs})$ and penalty functional $\mathcal{R}(f)$ are given by the quadratic norm powers, i.e.,*

$$\mathcal{S}(g, g^{obs}) = \frac{1}{2}\left\|g - g^{obs}\right\|_\mathbb{Y}^2 \quad and \quad \mathcal{R}(f) = \frac{1}{2}\|f - f_0\|_\mathbb{X}^2.$$

If the range condition (3.36) holds true with $\Psi(\lambda) = (-\ln(\lambda))^{-p}$ for $p > 0$ and a upper bound $\|g^\dagger - g^{obs}\|_\mathbb{Y} \leq \delta$ where the parameter $\alpha > 0$ is chosen as $\delta^2 = \alpha(-\ln(\lambda))^{-2p}$, then

we obtain the following convergence rates

$$\left\|\widehat{f}_{\alpha^*} - f^\dagger\right\|_{\mathbb{X}} = O\left((-\ln(\delta))^{-p}\right) \tag{3.53}$$

for all $\delta \to 0$.

Proof. See Corollary 3.35 [94]. $\qquad\square$

3.3.2 Alternative proofs of convergence rates

We already discussed in detail several advantages of variational regularization methods in section 3.2.3. To verify the feasibility of the variational regularization methods for linear inverse problems with additive random noise, we will further study the convergence analysis for our proposed methods and give alternative proofs of convergence rates for variational regularization of ill-posed inverse problems based on the deterministic convergence results that we derived in section 3.3.1.

Now we present the first main result of our thesis on the convergence rates for variational regularization with additive random noise, where we use variational source conditions (3.38) in the quadratic Hilbert spaces.

Theorem 3.24 (Convergence rates under general source conditions). *Suppose that Assumptions 3.1 and 3.4 hold true. Let Assumption 3.18 be satisfied with the variational inequality (3.38) and the Tikhonov functional (3.27) has a global minimizer $\widehat{f}_\alpha = R_\alpha(Tf^\dagger + \varepsilon W)$. Then the reconstruction error in expectation is bounded by*

$$\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|^2\right] \le 4\widetilde{C}\varepsilon^2 \frac{\widetilde{\Gamma}(\alpha)}{\alpha} + 4(-\Phi)^*\left(-\frac{1}{2\alpha}\right) \tag{3.54}$$

for all $\alpha > 0$, where $\widetilde{C} > 0$ given in Theorem 3.12. With the a-priori parameter choice rule

$$\alpha^* = \left(\frac{\Theta}{\widetilde{\Gamma}}\right)^{-1}(\varepsilon^2), \tag{3.55}$$

where $\Theta(\alpha) = \alpha(-\Phi)^*\left(-\frac{1}{2\alpha}\right)$, we obtain the following convergence rate

$$\sqrt{\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2\right]} \le 2(2C_\beta + 1)\widetilde{\Gamma}\left(\left(\frac{\Theta}{\widetilde{\Gamma}}\right)^{-1}(\varepsilon^2)\right). \tag{3.56}$$

Proof. Since $\widehat{f}_\alpha$ in (3.27) is a minimizer of the Tikhonov functional, we have the following inequality

$$\frac{1}{2\alpha}\|T\widehat{f}_\alpha - g^{\text{obs}}\|_{\mathbb{Y}}^2 + \frac{1}{2}\|\widehat{f}_\alpha\|^2 \le \frac{1}{2\alpha}\|Tf^\dagger - g^{\text{obs}}\|_{\mathbb{Y}}^2 + \frac{1}{2}\|f^\dagger\|_{\mathbb{X}}^2.$$

By Assumption 3.18 with inequality (3.38) and $\beta = \frac{1}{2}$, we have

$$\begin{aligned}
\frac{1}{4}\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^2 &\le \frac{1}{2}\|\widehat{f}_\alpha\|_{\mathbb{X}}^2 - \frac{1}{2}\|f^\dagger\|_{\mathbb{X}}^2 + \Phi\left(\|T\widehat{f}_\alpha - Tf^\dagger\|_{\mathbb{Y}}^2\right) \\
&\le \frac{1}{2\alpha}\left[\|Tf^\dagger - g^{\text{obs}}\|_{\mathbb{Y}}^2 - \|T\widehat{f}_\alpha - g^{\text{obs}}\|_{\mathbb{Y}}^2\right] + \Phi\left(\|T\widehat{f}_\alpha - Tf^\dagger\|_{\mathbb{Y}}^2\right) \\
&\le \frac{1}{\alpha}\langle g^{\text{obs}} - g^\dagger, T\widehat{f}_\alpha - g^\dagger\rangle - \frac{1}{2\alpha}\|T\widehat{f}_\alpha - Tf^\dagger\|_{\mathbb{Y}}^2 + \Phi\left(\|T\widehat{f}_\alpha - Tf^\dagger\|_{\mathbb{Y}}^2\right)
\end{aligned}$$

$$\leq \frac{\mathbf{err}(T\widehat{f}_\alpha)}{\alpha} + \sup_{\tau \geq 0}\left[\frac{1}{-2\alpha}\tau - (-\Phi)(\tau)\right]$$

$$= \frac{\mathbf{err}(T\widehat{f}_\alpha)}{\alpha} + (-\Phi)^*\left(-\frac{1}{2\alpha}\right). \tag{3.57}$$

By taking expectations on the both side of (3.57) and using the result of Theorem 3.12, we have

$$\mathbf{E}\left[\|\widehat{f}_\alpha - f^\dagger\|^2\right] \leq \frac{4\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right]}{\alpha} + 4(-\Phi)^*\left(-\frac{1}{2\alpha}\right)$$

$$\leq 4\widetilde{C}\varepsilon^2\frac{\widetilde{\Gamma}(\alpha)}{\alpha} + 4(-\Phi)^*\left(-\frac{1}{2\alpha}\right).$$

To show (3.54), the parameter choice rule (3.55) imples the equation

$$\frac{\widetilde{\Gamma}(\alpha)}{\alpha} = (-\Phi)^*\left(-\frac{1}{2\alpha}\right).$$

Finally by using the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$, we obtain (3.56). $\square$

Comparing the rates (3.19) in Theorem 3.5 with the convergence rates (3.56) in Theorem 3.24 we could also achieve order optimality of the varitional regularization for the additive random noise under variational source conditions. As a consequence of the above theorem, we can prove optimal convergence rates of the variational regularization for mildly ill-posed inverse problems with the additive random noise.

Corollary 3.25 (Optimal rates for Hölder-type source conditions). *Suppose that Assumptions 3.1 and 3.4 hold true. Let Assumption 3.18 be satisfied with the variational inequality (3.38) and the index function Φ satisfies $\Phi(\alpha) = \beta\alpha^{\frac{2\nu}{2\nu+1}}$. Let $\widehat{f}_\alpha = R_\alpha(Tf^\dagger + \varepsilon W)$ be a global minimizer. Then we have*

$$\mathbf{E}\left[\|\widehat{f}_\alpha - f^\dagger\|^2\right] \leq 4\widetilde{C}\varepsilon^2\alpha^{-1-\frac{1}{2p}} + 4^\nu\bar{C}\alpha^{2\nu} \tag{3.58}$$

for all $\alpha > 0$, where $\bar{C} = \left[-\frac{2\nu+1}{\nu} + \frac{1}{2}\left(\frac{2u+1}{\nu}\right)^{\frac{2\nu}{2\nu+1}}\right]$. With the a-priori parameter choice rule

$$\alpha^* = \varepsilon^{\frac{4}{4\nu+2+1/p}}, \tag{3.59}$$

we obtain the following convergence rate

$$\sqrt{\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2\right]} \leq C\varepsilon^{\frac{4\nu}{4\nu+2+1/p}}, \tag{3.60}$$

where $C = \sqrt{\left(4^\nu\bar{C} + 4\widetilde{C}\right)}.$

Proof. By Lemma 3.20 and replace γ by $-\frac{1}{2\alpha}$, we obtain

$$(-\Phi)^*\left(-\frac{1}{2\alpha}\right) = \bar{C}\left(-\frac{1}{2\alpha}\right)^{-2\nu} = 4^\nu\bar{C}\alpha^{2\nu}. \tag{3.61}$$

Now we take an expectation on the both side of (3.57) and together with (3.61) and Corollary 3.13, we obtain

$$
\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2\right] \leq \frac{4\mathbf{E}\left[\mathbf{err}(T\widehat{f}_\alpha)\right]}{\alpha} + 4^\nu \bar{C}\alpha^{2\nu} \leq 4\widetilde{C}\varepsilon^2 \alpha^{-1-\frac{1}{2p}} + 4^\nu \bar{C}\alpha^{2\nu}.
$$

This proves the assertion (3.58). To show (3.60), the parameter choice rule (3.59) imples $\varepsilon^2\alpha^{-1-\frac{1}{2p}} \leq \alpha^{2\nu}$. Finally, by using the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, we obtain (3.60). $\qquad\square$

Analogously, by comparing the rates in Corollary 3.13 and 3.25, we find that the convergence rate that we obtained in Corollary 3.25 is same as the rate of convergence by using spectral regularization methods. On the other hand, we can show in the similar way optimal convergence rates in expectation for variational regularization under logarithmic source conditions, which is shown as follows.

Corollary 3.26 (Optimal convergence rates for logarithmic source conditions). *Suppose that Assumptions 3.1 and 3.4 hold true. Let Assumption 3.18 be satisfied with the form of inequality (3.38) and the index function Φ satisfy $\Phi(\alpha) = \beta(-\ln\lambda)^{-2p}$. Let $\widehat{f}_\alpha = R_\alpha(Tf^\dagger + \varepsilon W)$ be a global minimizer of Tikhonov functional (3.27). Then we have*

$$
\mathbf{E}\left[\|\widehat{f}_\alpha - f^\dagger\|^2\right] \leq \frac{4c\widetilde{C}}{\alpha}\varepsilon^2\left(-\frac{1}{2\bar{t}}\ln\alpha\right)^{d/2} + 4(-\ln\alpha)^{-2p}(1 + o(1)) \tag{3.62}
$$

for all $\alpha > 0$, where $\widetilde{C} > 0$. With the a-priori parameter choice rule

$$
\alpha^* = (-\ln\varepsilon)^{2p}(1 + o(1)), \tag{3.63}
$$

we obtain the following convergence rate

$$
\sqrt{\mathbf{E}\left[\|\widehat{f}_{\alpha^*} - f^\dagger\|_{\mathbb{X}}^2\right]} = O((-\ln\varepsilon)^{-p}), \tag{3.64}
$$

as $\varepsilon \to 0$.

Proof. By (3.54) of Theorem 3.24, we can replace $\widetilde{\Gamma}$ by $\widetilde{\Gamma}(\alpha) = c\left(-\frac{1}{2\bar{t}}\ln\alpha\right)^{d/2}$ and the convex conjugation of the index function Φ by (3.47) such that we obtain (3.62). To show (3.64), the parameter choice rule (3.63) implies

$$
\alpha^{-1}\left(-\frac{1}{2\bar{t}}\ln\alpha\right)^{d/2} \leq (-\ln\alpha)^{-2p}(1 + o(1))
$$

and using the inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$, then we derive (3.64). $\qquad\square$

One of the main reasons to give the alternative proofs of convergence rates for variational regularization is to present the advantages of variational methods in Chapters 5, 6 and 7. As we seen in previous proofs, we still study the convergence analysis for variational regularization methods of linear inverse problems based on the spectral theory without applying concentration inequalities to bound the propagated noise error. In latter chapters, we focus mainly on the variational regularization method of nonlinear inverse problems in different scenario with Poisson and empirical process data and improve the reconstruction error in expectation, where the data fidelity functional $\mathcal{S}$ is given by Kullback-Leibler divergence. To show convergence rates for variational methods of inverse problems with non-Gaussian noise

error, we mainly propose to establish deviation inequalities for Poisson and empirical processes. In such cases regularization methods based on the spectral theory no longer applicable.

Chapter 4

Deviation inequalities for stochastic processes

The main purpose of this chapter is to deal with deviation inequalities for stochastic processes in Besov spaces with negative smoothness indexes, which play an important role in a study of the consistency and the rates of convergence of estimators. Since the derivation of optimal convergence rates for variational regularization methods with stochastic process data such as Poisson and empirical processes data is dependent crucially on deviation or concentration inequalities, with the result of this chapter we could improve error bounds in expectation as we will discuss in Chapter 5 and Chapter 7. To derive such deviation inequalities in Besov spaces, we will start with an overview of definitions and basic properties of Fourier and wavelet-based Besov spaces.

In section 4.1 we will give precise definitions of Besov spaces and recall several important properties of these spaces that we need in our thesis. In section 4.2 we will derive in particular deviation inequalities for Poisson and empirical processes in Fourier Besov spaces based on the concentration inequalities in Lemma 2.4 and Lemma 2.9, respectively. The main difficulties in our analysis are to bound the expectations for the negative normed Besov spaces of Poisson and empirical noise processes. Although the deviation inequality for Gaussian white noise has been derived by Veraar [92], similar results for Poisson and empirical process have not studied so far. Therefore, we will mainly discuss these deviation inequalities for Poisson and empirical processes in negative Besov norms.

4.1 Besov spaces and their properties

In this section we propose a well-known class of function spaces, namely, call them Besov spaces. First of all, we will give two equivalent definitions of the Besov spaces, one is based on a dyadic resolution of unity in the Fourier domain and another one is based on wavelet decompositions. In particular, we focus mainly on the derivation of deviation inequalities for stochastic processes in Fourier based periodic Besov spaces.

4.1.1 Besov spaces

There exist several equivalent definitions on Besov space, which is a useful function space to measure regularity properties of functions. To understand the chapter more readable, we first introduce a definition of Besov space in the Fourier domain $\mathbb{R}^d$. Let $\{\varphi_l(t)\}_{l=0}^{\infty}$ be a smooth dyadic resolution of unity in $\mathbb{R}^d$, then we have

- $\mathrm{supp}\,\varphi_0 \subset \{t \in \mathbb{R}^d : |t| \leq 2\}$

- $\operatorname{supp}\varphi_l \subset \{t \in \mathbb{R}^d : 2^{l-1} \le |t| \le 2^{l+1}\} \quad l \in \mathbb{N}.$

- $\sup_{t \in \mathbb{R}, l \in \mathbb{N}_0} |\varphi_l(t)| < \infty \quad$ and $\quad \sum_{l=0}^{\infty} \varphi_l(t) = 1.$

Let $\mathscr{D}(\mathbb{R}^d)$ denote a Schwarz space and $\mathscr{D}'(\mathbb{R}^d)$ denote the space of all complex-valued tempered distributions on $\mathbb{R}^d$.

Definition 4.1 (**Besov space on domain $\mathbb{R}^d$**). *Let $s \in \mathbb{R}$, $p \in (1, \infty]$ and $q \in [1, \infty]$ be given parameters of smoothness, integrability and summability respectively. Then we define the periodic Besov space as*

$$B_{p,q}^s(\mathbb{R}^d) := \{P \in \mathscr{D}'(\mathbb{R}^d) : \|P\|_{B_{p,q}^s(\mathbb{R}^d)} < \infty\},$$

where the Besov norm of P is given by

$$\|P\|_{B_{p,q}^s(\mathbb{R}^d)} := \begin{cases} \left(\sum_{l \ge 0} 2^{slq} \|\mathcal{F}\varphi_l \mathcal{F}P\|_{L^p(\mathbb{R}^d)}^q\right)^{1/q}, & q < \infty, \\ \sup_{l \ge 0} 2^{sl} \|\mathcal{F}\varphi_l \mathcal{F}P\|_{L^p(\mathbb{R}^d)}, & q = \infty. \end{cases} \tag{4.1}$$

Since it is not our purpose to study the deviation inequality for stochastic processes in negative Besov spaces with the domain $\mathbb{R}^d$, we are particularly interested in investigating the problem of deviation inequality rather in a bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$. Therefore, we give in the following a precise definition for a Besov space with domain $\mathbb{M}$.

Definition 4.2 (**Besov space on Lipschitz domain $\mathbb{M}$**). *Let $s \in \mathbb{R}$, $p \in (1, \infty]$ and $q \in [1, \infty]$ be given parameters of smoothness, integrability and summability respectively. Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded Lipschitz domain and Besov spaces on $\mathbb{M}$ by*

$$B_{p,q}^s(\mathbb{M}) = \begin{cases} \{P \in \mathscr{D}'(\mathbb{M}) : P = \tilde{P}|_{\mathbb{M}}, \tilde{P} \in B_{p,q}^s(\mathbb{R}^d)\} & \text{if } s \ge 0 \\ \{P \in \mathscr{D}'(\mathbb{M}) : P = \tilde{P}|_{\mathbb{M}}, \tilde{P} \in B_{p,q}^s(\mathbb{R}^d), \operatorname{supp}\tilde{P} \subset \overline{\mathbb{M}}\} & \text{if } s < 0. \end{cases}$$

and

$$\|P\|_{B_{p,q}^s(\mathbb{M})} := \inf \|\tilde{P}\|_{B_{p,q}^s(\mathbb{R}^d)},$$

where the infimum is taken over all extensions $\tilde{P}$.

The Besov spaces $B_{p,q}^s$ are Banach spaces such that

$$\mathscr{D}(\mathbb{M}) \hookrightarrow B_{p,q}^s(\mathbb{M}) \hookrightarrow \mathscr{D}'(\mathbb{M}),$$

where the first embedding is dense if $p, q < \infty$. The theory of the function space $B_{p,q}^s(\mathbb{M})$ has been discussed in detail by Triebel [88, 89].

4.1.2 Wavelet basis

In this subsection we also introduce another equivalent type of norm on Besov spaces $B_{p,q}^s(\mathbb{R}^d)$ based on Wavelet decompositions. One of the reasons to introduce this type of Besov norm is to make our thesis more readable when we present the convergence rate results for Poisson inverse problems based on Wavelet decompositions studied by Antoniadis and Bigot [2], which will be restated in Chapter 5. For the general theory of wavelets in the function spaces we refer to [96].

For $h \in \mathbb{N}$, let $C^h(\mathbb{R}^d)$ denotes the space of h times continuously differentiable complex-valued functions on $\mathbb{R}^d$ and $|I| \le h$. Let $\psi_F \in C^h(\mathbb{R}^d)$ and $\psi_M \in C^h(\mathbb{R}^d)$ denote either the mother wavelet or the father wavelet of a d-dimensional wavelet basis of L^2. These wavelets are real and have compact support. Moreover,

$$\int_{\mathbb{R}^d} \psi_F(x)\mathrm{d}x = 1, \quad \int_{\mathbb{R}^d} x^I \psi_M(x)\mathrm{d}x = 0$$

for $I \in \mathbb{N}_0^d, |I| \le h$. Define $\psi_{j,k}$ by

$$\psi_{j,k}(x) = \begin{cases} \psi_F(x-k), & j = 0, k \in \mathbb{Z}^d, \\ 2^{\frac{jd}{2}} \psi_M(2^{j-1}x - k), & j \in \mathbb{N}_0, k \in \mathbb{Z}^d. \end{cases} \tag{4.2}$$

Then the sequence $\{\psi_{j,k}\}_{j\in\mathbb{N}_0, k\in\mathbb{Z}^d}$ in (4.2) is an orthonormal basis in $L^2(\mathbb{R}^d)$. To transform the wavelet basis of $L^2(\mathbb{R}^d)$ into a wavelet basis of $L^2(\mathbb{T}^d)$ with the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, we need to periodize each basis of $\{\psi_{j,k}\}_{j\in\mathbb{N}_0,k\in\mathbb{Z}^d}$. Let $l = 1, \cdots, E_j$ with

$$E_j := \begin{cases} 1, & \text{if } j = 0, \\ 2^d - 1, & \text{else.} \end{cases}$$

Then, one can replace ψ_F, ψ_M by

$$\psi_F^l(\cdot) = \psi^0(2^l \cdot), \quad \psi_M^l(\cdot) = \psi_l(2^l \cdot)$$

and $\psi_{j,k}^l = 2^{\frac{l}{2}} \psi_{j,k}(2^l \cdot)$. More precisely, $\psi_{j,k}$ is given by

$$\psi_{j,k}^l(x) = \begin{cases} \psi_F^0(x-k), & j = 0, l = 0, k \in \mathbb{Z}^d, \\ 2^{\frac{jd}{2}} \psi_M^l(2^{j-1}x - k), & j \in \mathbb{N}, l \in [1, 2^d - 1], k \in \mathbb{Z}^d, \end{cases}$$

Then all the functions $\psi_{j,k}^l$ have compact supports, i.e.,

$$\operatorname{supp} \psi_F^0 \subset \{x \in \mathbb{R}^d : |x| \le \frac{1}{2}\} \quad \operatorname{supp} \psi_M^l \subset \{x \in \mathbb{R}^d : |x| \le \frac{1}{2}\}.$$

Let

$$\mathbb{P}_j = \begin{cases} \{k \in \mathbb{Z}^d : 0 \le k \le 2^l - 1\}, & j = 0, \\ \{k \in \mathbb{Z}^d : 0 \le k \le 2^{j+l-1} - 1\}, & j \in \mathbb{N}. \end{cases}$$

Define $\psi_{j,k}^{l,per}$ on the d-dimensional torus $\mathbb{T}^d$ by

$$\psi_{j,k}^{l,per}(x) = \sum_{z\in\mathbb{Z}^d} \psi_{j,k}^{l,per}(x+z).$$

Then $\{\psi_{j,k}^{l,per}\}_{j\in\mathbb{N}_0,k\in\mathbb{P}_j}$ is an orthonormal basis in $L^2(\mathbb{T}^d)$. The scalar product on $\mathbb{T}^d$ is defined by

$$\langle f, \psi_{j,k}^{l,per} \rangle = \int_{\mathbb{T}^d} f(x)\psi_{j,k}^{l,per}(x)\mathrm{d}x.$$

Then for any $s \in \mathbb{R}$ and $p, q \in [1, \infty]$ with $h > |s|$, the Besov norm of a function f is defined by

$$\|P\|_{B_{p,q}^s} := \left(\sum_{j=1}^{\infty} 2^{jq\left(s+d\left(\frac{1}{2}-\frac{1}{p}\right)\right)} \left(\sum_{k \in \mathbb{P}_j} \sum_{l \in E_j} \left| \langle P, \psi_{j,k}^{l,per} \rangle \right|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \tag{4.3}$$

with usual modification if $p = \infty$ or $q = \infty$. More precisely, the Besov space $B_{p,q}^s(\mathbb{T}^d)$ consists of L^p functions with finite Besov norm for $s > 0$ and $p \in [1, \infty)$, however, it consists of continuous functions with finite Besov norm for $s > 0$ and $p = \infty$.

4.1.3   Some basic properties of Besov spaces

In this subsection we will collect several important properties of (homogeneous) Besov spaces that we needed in this thesis. Due to our particular interest, we do not consider the Besov space $B_{p,q}^s(\mathbb{R}^d)$ but investigate the deviation inequality for the processes in $B_{p,q}^s(\mathbb{M})$ for a bounded Lipschitz domain $\mathbb{M}$. Therefore, we will restrict ourselves with Besov space on the Lipschitz domain $\mathbb{M}$ in the rest of the thesis.

First of all, we recall some elementary embedding results for Besov spaces and Sobolev spaces. For any $s \in \mathbb{R}, \varepsilon > 0, p \in (1, \infty)$ and $1 \le q_1 \le q_2 \le \infty$, we have

$$B_{p,\infty}^{s+\varepsilon} \subset B_{p,1}^s \subset B_{p,q_1}^s \subset B_{p,q_2}^s \subset B_{p,\infty}^s \subset B_{p,\infty}^{s-\varepsilon}, \tag{4.4}$$

where "$\subset$" means continuous embedding, i.e., if $B_{p,q_1}^s \subset B_{p,q_2}^s$, then for any $P \in \mathscr{D}'(\mathbb{M})$ we have $\|P\|_{B_{p,q_1}^s} \le C\|P\|_{B_{p,q_2}^s}$ with constant $C > 0$. There exist some important embedding results between Besov and Sobolev spaces. For any $s \in \mathbb{R}$, one has the following results

$$B_{p,2}^s(\mathbb{M}) \subset H_p^s(\mathbb{M}) \subset B_{p,p}^s(\mathbb{M}), \quad \text{if } p \in [2, \infty),$$
$$B_{p,p}^s(\mathbb{M}) \subset H_p^s(\mathbb{M}) \subset B_{p,2}^s(\mathbb{M}), \quad \text{if } p \in (1, 2)$$

and for any $\varepsilon > 0$ and $r, s \in [1, \infty]$, we have

$$B_{p,q}^{s+\varepsilon}(\mathbb{M}) \subset H_p^{s,r}(\mathbb{M}) \subset B_{p,q}^{s-\varepsilon}(\mathbb{M}).$$

Note that $B_{2,2}^s(\mathbb{M}) = H_2^s(\mathbb{M})$ and $B_{2,2}^0(\mathbb{M}) = L^2(\mathbb{M})$, but these special equivalences do not hold for $p \ne 2$. For more details of embedding results, we refer to [88].

In addition, we could also introduce embedding results for different values of integrability. For any $1 \le p_1 \le p_2 < \infty, 1 \le q_1 \le q_2 < \infty$ and $-\infty < s_2 \le s_1 < \infty$, the embedding

$$B_{p_1,q_1}^{s_1}(\mathbb{M}) \subset B_{p_2,q_2}^{s_2}(\mathbb{M}) \tag{4.5}$$

is continuous if $s_1 - \frac{d}{p_1} \ge s_2 - \frac{d}{p_2}$ and compact if $s_1 > s_2$. Let $1 \le p_1 \le p_2 < \infty$, $q \in [1, \infty)$ and $s \in \mathbb{R}$, then

$$B_{p_1,q}^{s_1}(\mathbb{M}) \subset B_{p_2,q}^{s_2}(\mathbb{M}),$$

see Remark 1 in paragraph 4.3.3 [89]. Moreover, if the smoothness index $s \in (0, 1)$ is not an integer number, then $B_{p,p}^s = W^{s,p}$ is the Sobolev space with norm

$$\|P\|_{W^{s,p}(\mathbb{R}^d)} = \|P\|_{L^p(\mathbb{R}^d)} + \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \frac{|P(x) - P(y)|^p}{|x-y|^{d+sp}} dx dy \right)^{\frac{1}{p}}.$$

In particular, Sobolev spaces with $p = 2$ are equivalent to H^s, i.e., $W^{s,2} = H^s$, which form a Hilbert space.

Definition 4.3 (Multiplication algebra). *A Besov space $B_{p,q}^s(\mathbb{M})$ is said to be a multiplication algebra if for all $P_1, P_2 \in \mathbb{B}_{p,q}^s(\mathbb{M})$, the product $P_1 \cdot P_2$ exists and belongs to the space $B_{p,q}^s(\mathbb{M})$ and if there exists a positive constant c such that*

$$\|P_1 P_2\|_{B_{p,q}^s(\mathbb{M})} \leq c \|P_1\|_{B_{p,q}^s(\mathbb{M})} \|P_2\|_{B_{p,q}^s(\mathbb{M})} \tag{4.6}$$

holds for all $P_1, P_2 \in B_{p,q}^s(\mathbb{M})$.

Theorem 4.4. *Let $s \in \mathbb{R}$ and $p, q \in [1, \infty]$. Then the following three statements are equivalent:*

(i) *$B_{p,q}^s(\mathbb{M})$ is a multiplication algebra.*

(ii) *$B_{p,q}^s(\mathbb{M}) \hookrightarrow C(\mathbb{M})$, where $C(\mathbb{M})$ is the space of all bounded and uniformly continuous functions on the bounded domain $\mathbb{M}$.*

(iii) *There holds either for $p, q \in [1, \infty]$ and $s > \frac{d}{p}$ or $p \in [1, \infty]$, $q = (0, 1]$ and $s = \frac{d}{p}$.*

Proof. The above statements follow from the Theorem 2.8.3 in [89] if $\mathbb{M} = \mathbb{T}^d$. For the bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$, we use Definition 4.1. $\square$

We now recall a main theorem concerning interpolations of Besov spaces $B_{p,q}^s$ that we need in Chapter 5. For proofs, see Theorem 2.4.1 in Triebel [88]:

Theorem 4.5 (Interpolation inequality). *Let $s_0, s_1 \in \mathbb{R}, s_0 \neq s_1, p \in (1, \infty), q_0, q_1, q \in [1, \infty]$ and $\theta \in (0, 1)$. Let*

$$B_{p,q}^s(\mathbb{M}) = (B_{p,q_0}^s(\mathbb{M}), B_{p,q_1}^s(\mathbb{M}))_{\theta,q}.$$

Then there exists a constant $C_B > 0$ such that the following interpolation inequality holds true

$$\|P\|_{B_{p,q}^s(\mathbb{M})} \leq C_B \|P\|_{B_{p,q_0}^s(\mathbb{M})}^{\theta} \|P\|_{B_{p,q_1}^s(\mathbb{M})}^{1-\theta} \tag{4.7}$$

for all $P \in B_{p,q_0}^{s_0}(\mathbb{M}) \cap B_{p,q_1}^{s_1}(\mathbb{M})$.

Proof. For $\mathbb{M} \subset \mathbb{R}^d$, the statement follows from Definition 4.2 and Theorem 2.4.1 (a) in Triebel [88]. $\square$

Theorem 4.6 (Duality). *Let $s \in \mathbb{R}, p \in [1, \infty]$ and $q \in [1, \infty)$. Then*

$$\left(B_{p,q}^s(\mathbb{M})\right)' = B_{p',q'}^{-s}(\mathbb{M}), \quad \frac{1}{p} + \frac{1}{p'} = \frac{1}{q} + \frac{1}{q'} = 1. \tag{4.8}$$

Proof. See Theorem 2.6.1 [88]. $\square$

As a consequence of Theorem 4.6, there exists a constant $C_H > 0$ such that the Hölder inequality

$$\|P\|_{L^2(\mathbb{M})} \leq C_H \|P\|_{B_{p_1,q_1}^s(\mathbb{M})} \|P\|_{B_{p_2,q_2}^{-s}(\mathbb{M})} \tag{4.9}$$

holds for all $P \in B_{p_1,q_1}^{s_1}(\mathbb{M}) \cap B_{p,q_2}^{s_2}(\mathbb{M})$ with $1/p_1 + 1/p_2 = 1$ and $1/q_1 + 1/q_2 = 1$.

4.2   Deviation inequalities

As we described previously, large deviation inequality for Gaussian white noise has been derived by Veraar [92]. For a Gaussian white noise process,

$$\mathbf{E}\left[\|W\|_{B_{p,\infty}^{-d/2}(\mathbb{M})}\right] = \mathbf{E}\left[\sup_{l\geq 0} 2^{-\frac{d}{2}l}\|W_l\|_{L^2(\mathbb{M})}\right] < \infty \qquad (4.10)$$

holds true with $s = d/2$ and $p \in [1,\infty)$, see Proposition 2.3 [92]. In the following, we will recall this deviation inequality for Gaussian white noise in negative indexed Besov norms.

Theorem 4.7 (**Deviation inequality for Gaussian process**). *Let W denote a Gaussian white noise process in $L^2(\mathbb{M})$ and let $p \in [1,\infty)$. Then there are constants $M, r > 0$ depending on p and the dimension d such that*

$$\mathbf{P}\left(\left|\|W\|_{B_{p,\infty}^{-d/2}(\mathbb{M})} - M\right| \geq r\right) \leq \exp\left(-\frac{r^2}{4\vartheta^2}\right) \qquad (4.11)$$

holds for any $b \geq 0$. The constant M can be taken to be the median of $\|W\|_{B_{p,\infty}(\mathbb{M})}$ and σ can be chosen as

$$\vartheta = \begin{cases} (2\pi)^{\frac{d}{p}-\frac{d}{2}} & \text{for } p \in [1,2), \\ 2^{\frac{3d}{2}-\frac{3d}{p}} & \text{for } p \in [2,\infty). \end{cases}$$

Proof. See Corollary 3.7 in Veraar [92].   □

4.2.1   Bounds on Besov norms of empirical and Poisson processes

In order to establish large deviation inequalities for Poisson and empirical processes in negative Besov norms based on Fourier analysis, we need to introduce some notations. In the following we study Poisson and empirical noise processes and we denote them by $P = G_n - g^\dagger$, where G_n is either an empirical process or a rescaled Poisson process with intensity $g^\dagger$. We define

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

Note that we restrict ourselves here noise processes $G_n - g^\dagger$ in the measurement manifold $\mathbb{M}$ throughout the thesis. To apply Lemma 2.9 for Poisson processes and Lemma 2.4 for the empirical processes to each P_l respectively, we introduce a new set

$$\mathcal{S}_l := \{\mathbf{F} \in L^2(\mathbb{M}) : \|\mathbf{F}\|_{L^2} \leq 2^{-sl} \text{ and } \widehat{\mathbf{F}}(\kappa) = 0 \text{ if } \kappa \notin \text{supp } \phi_l\}. \qquad (4.12)$$

This gives

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

$$= \sup_{\|\mathbf{F}\|_{L^2}\leq 2^{-sl}} \left|\sum_{\kappa\in\mathbb{Z}^d} \widehat{\mathbf{F}}(\kappa)\widehat{P_l}(\kappa)\right|$$

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

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

$$\leq \sup_{\mathbf{F}\in\mathcal{S}_l} \left|\int_{\mathbb{M}} \mathbf{F}(x)(\mathrm{d}G_n - g^\dagger \mathrm{d}x)\right|. \tag{4.13}$$

4.2.2 Deviation inequalities in negative Besov norms

In this subsection we will establish the deviation inequalities for Poisson and empirical processes based on the results obtained by Reynaud-Bouret [76] and Massart [65], respectively. Deviation inequalities for Poisson and empirical processes are crucial tool for studying convergence analysis of variational regularization with Poisson and empirical process data in later chapters. We now need to bound the expectation of $Z^{(s)}$ in Besov spaces with negative smoothness index $s > d/2$. These results are important to show the deviation inequalities for the stochastic processes in negative Besov norms.

Now we are able to control the expectation of $Z^{(s)}$ for Poisson and empirical processes.

Theorem 4.8 (**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 (4.13) with the smoothness index $s > \frac{d}{2}$. Then there exists a constant C such that*

$$\mathbf{E}\left[\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^{-s}(\mathbb{M})}\right] < \frac{CD}{\sqrt{n}}\|g^\dagger\|_{L^1(\mathbb{M})}. \tag{4.14}$$

Proof. Using the continuous embedding (4.4) we know that $B_{2,2}^{-s} \subset B_{2,\infty}^{-s}$ for $s > d/2$, then

$$\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^{-s}(\mathbb{M})} \leq C_1 \left\|G_n - g^\dagger\right\|_{B_{2,2}^{-s}(\mathbb{M})} \tag{4.15}$$

holds true for some constant $C_1 > 0$. From Lemma A.3 in Werner and Hohage [95], we have

$$\mathbf{E}\left[\left\|G_n - g^\dagger\right\|_{H^{-s}(\mathbb{M})}\right] \leq \frac{C_2 D}{\sqrt{n}}\|g^\dagger\|_{L^1(\mathbb{M})} \tag{4.16}$$

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 (4.15), we immediately derive (4.14) with $C = C_1 C_2 > 0$. $\qquad\square$

As we see from Theorem 4.8, we could show that

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

for fixed n if $s > d/2$. As for the limiting case $s = d/2$, we will give a conjecture for the expectation of noise processes $G_n - g^\dagger$ in negative Besov norm up to logarithmic factor.

Conjecture 4.9. *For an empirical or a Poisson process G_n with intensity $g^\dagger$ and the expected count n, we have*

$$\sup_{n\in\mathbb{N}} \mathbf{E}\left[\frac{\sqrt{n}}{(\ln n)^p}\|G_n - g^\dagger\|_{B_{2,\infty}^{-d/2}(\mathbb{M})}\right] < \infty \tag{4.17}$$

for $p \geq 0$.

Remark 4.10. *In this conjecture, no smaller value of s is possible as Poisson and empirical distributions $G_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{x_j}$ do not belong to $B_{2,\infty}^{-s}(\mathbb{M})$ for $s < d/2$, however $\left\|\delta_{x_j}\right\|_{B_{2,\infty}^{-d/2}(\mathbb{M})}$ is uniformly bounded in x_j. As a consequence, we can show by the triangle inequality that*

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

In this situation only the exact rate of decay is unknown.

This conjecture is crucial to improve our reconstruction error bound between the regularized estimator and the true solution. If one can prove this conjecture in future we can achieve our purpose of the thesis and derive optimal rates of convergence for variational regularization and iteratively regularized Newton-type methods as we study in Section 5.2.2 and 7.3, respectively. The last result of this chapter is to derive deviation inequalities for Poisson and empirical processes in negative Besov norms.

Theorem 4.11 (Deviation inequality for empirical processes). *Suppose that $\mathbb{M} \subset \mathbb{R}^d$ is a bounded d-dimensional Lipschitz manifold. Let G_n be an empirical process with intensity $g^\dagger$ in Besov norms of $B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})$ with $\tilde{d} > d$. Then the following deviation inequality*

$$\mathbf{P}\left(\sqrt{n}\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}} \geq C_\rho + \eta\right) \leq \exp\left(-\frac{\eta}{c_1}\right) \tag{4.18}$$

holds true for all $n, \eta \geq c_1$, where

$$c_1 = \sqrt{8} + 34.5 \quad and \quad C_\rho = 2CD.$$

Moreover, if Conjecture 4.9 holds true with $\tilde{d} = d$ such that there exists a constant $C_p > 0$, then

$$\mathbf{P}\left(\sqrt{n}\|G_n - g^\dagger\|_{B_{2,\infty}^{-d/2}} \geq (C_p + \eta)(\ln n)^p\right) \leq \exp\left(-\frac{\eta}{c_1}\right) \tag{4.19}$$

for $p \geq 0$ and $n, \eta \geq c_1$.

Proof. The proof of the theorem can be established based on the concentration inequalities of Lemma 2.4, and we already bound the expectation of $Z^{(s)}$. Thus, before we apply Lemma 2.4, it is necessary to control the constants v_0 and b in (2.16). Since $Z^{(s)} = \sup_{l\geq 0} 2^{-sl}\|P_l\|_{L^2}$, we can control b by

$$b = \sup_{l\geq 0}\sup_{\mathbf{F}\in\mathcal{S}_l}\|\mathbf{F}\|_\infty$$
$$\leq \sup_{l\geq 0}\sup_{\mathbf{F}\in\mathcal{S}_l} 2^{\frac{d}{2}l}\|\mathbf{F}\|_{L^2}$$
$$\leq \sup_{l\geq 0} 2^{-(s-\frac{d}{2})l} \leq 1, \tag{4.20}$$

To control the constants v, we use Hölder's inequality

$$v = n\sup_{l\geq 0}\sup_{\mathbf{F}\in\mathcal{S}_l}\int_{\mathbb{M}}\mathbf{F}^2(x)g^\dagger(x)\mathrm{d}x$$
$$\leq n\sup_{l\geq 0}\|\mathbf{F}^2\|_{L^\infty(\mathbb{M})}\|g^\dagger\|_{L^1(\mathbb{M})}$$
$$\leq Cn\sup_{l\geq 0} 2^{dl}\|\mathbf{F}\|_{L^2(\mathbb{M})}^2\|g^\dagger\|_{L^1(\mathbb{M})}$$

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

$$\leq Cn \|g^\dagger\|_{L^1(\mathbb{M})}. \tag{4.21}$$

for some constant $C > 0$. We now apply Lemma 2.4 to $Z^{(s)}$:

$$\mathbf{P}\left[Z^{(s)} \geq (1+\varepsilon)\mathbf{E}[Z^{(s)}] + \frac{\sqrt{8\widehat{\eta}v}}{n} + \frac{\widehat{\eta}\mu_2(\varepsilon)b}{n} \right] \leq \exp(-\widehat{\eta}) \tag{4.22}$$

for all $\widehat{\eta} > 0$. By plugging (4.14), (4.20) and (4.21) into (4.22) and choosing $\varepsilon = 1$ as well as using $\mathbf{P}(A) = 1 - \mathbf{P}^c(A)$ with $\|g^\dagger\|_{L^1} = 1$, we obtain

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

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

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

for all $\widehat{\eta}, n > 0$. Letting

$$\eta = \left(\sqrt{8} + 34.5\right)\widehat{\eta} \quad \text{and} \quad C_\rho = 2CD$$

we derive

$$\mathbf{P}\left(Z^{(s)} \geq \frac{C_\rho + \eta}{\sqrt{n}} \right) \leq \exp\left(-\frac{\eta}{\sqrt{8} + 34.5} \right).$$

By taking $c_1 = \sqrt{8} + 34.5$, which yields (4.20).

If Conjecture 4.9 holds true with $\tilde{d} = d$ such that there exists a constant $C_p > 0$, then

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

for $p \geq 0$. By inserting above inequality into (4.22) and plugging (4.20) and (4.21), we have

$$\mathbf{P}\left[\|G_n - g\|_{B_{2,\infty}^{-d/2}(\mathbb{M})} \leq C_p \frac{(\ln n)^p}{\sqrt{n}} + \frac{\sqrt{8\widehat{\eta}}}{\sqrt{n}} + \frac{34.5}{n}\widehat{\eta} \right] \geq 1 - \exp(-\widehat{\eta})$$

Analogously, we also derive

$$\mathbf{P}\left(\|G_n - g\|_{B_{2,\infty}^{-d/2}(\mathbb{M})} \geq \frac{C_p(\ln n)^p + \eta}{\sqrt{n}} \right) \leq \exp\left(-\frac{\eta}{c_1} \right), \tag{4.23}$$

which completes the proof of theorem. $\square$

Theorem 4.12 (Deviation inequality for Poisson processes). *Assume that $\mathbb{M} \subset \mathbb{R}^d$ is a bounded d-dimensional Lipschitz domain. Let G_n be a normalized Poisson process with intensity $g^\dagger$ in Besov norms of $B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})$ with $\tilde{d} > d$. Then the following inequality*

$$\mathbf{P}\left(\sqrt{n}\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}} \geq C_\rho + \eta \right) \leq \exp\left(-\frac{\eta}{c_2} \right) \tag{4.24}$$

holds true for all $n, \eta \geq c_2$, where $c_2 = \sqrt{12} + 33.25$ and $C_\rho = 2CD$. Moreover, if Conjecture 4.9 holds true with $\tilde{d} = d$, then

$$\mathbf{P}\left(\sqrt{n}\|G_n - g^\dagger\|_{B_{2,\infty}^{-d/2}} \geq (C_p + \eta)(\ln n)^p\right) \leq \exp\left(-\frac{\eta}{c_2}\right) \qquad (4.25)$$

for $p \geq 0$ and $n, \eta \geq c_2$.

Proof. From the proof of Theorem 4.11, we already bound b and v such that it suffice to apply Lemma 2.9 to $Z^{(s)}$. Let $P = G_n - g^\dagger$ be Poisson noise process and $Z^{(s)} = \|P\|_{B_{2,\infty}^{-d/2}(\mathbb{M})}$. Then

$$\mathbf{P}\left[Z^{(s)} \geq (1+\varepsilon)\mathbf{E}[Z^{(s)}] + \frac{\sqrt{12\widehat{\eta}v}}{n} + \frac{\widehat{\eta}\mu_1(\varepsilon)b}{n}\right] \leq \exp(-\widehat{\eta}) \qquad (4.26)$$

for all $\widehat{\eta} > 0$. By inserting (4.14), (4.20) and (4.21) into (4.26) and choosing $\varepsilon = 1$ and using $\mathbf{P}(A) = 1 - \mathbf{P}^c(A)$ and $\|g^\dagger\|_{L^1} = 1$, we get

$$\mathbf{P}\left[Z^{(s)} \leq \frac{2CD}{\sqrt{n}} + \frac{\sqrt{12\widehat{\eta}}}{\sqrt{n}} + \frac{33.25}{n}\widehat{\eta}\right] \geq 1 - \exp(-\widehat{\eta})$$

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

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

for all $\widehat{\eta}, n > 0$. Setting $\eta = \left(\sqrt{12} + 33.25\right)\widehat{\eta}$ and $C_\rho = 2CD$, we derive

$$\mathbf{P}\left(Z^{(s)} \geq \frac{C_\rho + \eta}{\sqrt{n}}\right) \leq \exp\left(-\frac{\eta}{\sqrt{12} + 33.25}\right).$$

Finally, by taking $c_2 = \sqrt{12} + 33.25$, we obtain (4.24). Similarly, if Conjecture 4.9 holds true with $\tilde{d} = d$ and for $p \geq 0$, then (4.25) can be immediately obtained as the proof of Theorem 4.11. $\qquad \square$

By comparing the deviation inequality (4.11) for Gaussian white noise to (4.18) and (4.24), we find that (4.11) is available for the limiting case $s = \frac{d}{2}$, which means that it exhibits even faster decay of the tails rather than the cases of Poisson and empirical processes.

Chapter 5

Tikhonov-type regularization

In this chapter we study statistical convergence rate analysis for Tikhonov-type regularization for the solution of a (possibly nonlinear) ill-posed inverse problem described by a nonlinear equation

$$F(f^\dagger) = g^\dagger. \tag{5.1}$$

Here $F : \mathcal{B} \subset \mathbb{X} \to \mathbb{Y} = L^2(\mathbb{M})$ is the forward operator, $f^\dagger \in \mathcal{B}$ is an unknown quantity of object, $g^\dagger \in \mathbb{Y}$ is a non-negative, integrable function on some measurement manifold $\mathbb{M} \in \mathbb{R}^d$, which can be seen as exact data and $f^\dagger \in \mathbb{X}$ is unique true solution to the inverse problem.

To construct a stable approximation to $f^\dagger$ of (5.1), we study a Tikhonov-type estimator with a quadratic penalty term

$$\widehat{f}_\alpha \in \operatorname*{argmin}_{f \in \mathbb{X}} \left[\mathcal{S}(G_n, F(f)) + \frac{\alpha}{2} \|f\|_{\mathbb{X}}^2 \right], \tag{5.2}$$

where $\alpha > 0$ is regularization parameter, $\mathcal{S}$ is a Kullback-Leibler type data fidelity functional derived in Chapter 2 and $\mathcal{R}(f) = \frac{1}{2}\|f\|_{\mathbb{X}}^2$ is a quadratic penalty term. If the data fidelity term $\mathcal{S}(G_n, \cdot)$ and penalty term $\mathcal{R}$ are given by quadratic norm powers, then (5.2) is a standard Tikhonov regularization, which is already discussed in Chapter 3 for the case of Gaussian white noise. As we discussed in Chapter 3, the minimizer (3.28) can be derived by an exact computation for linear operators. Unfortunately, the minimizer $\widehat{f}_\alpha$ in (5.2) is not easy to compute due to the non-quadratic data fidelity functional $\mathcal{S}$ and non-linear operator F. Nevertheless, (5.2) is still a convex optimization problem for linear operators F. The minimization problem (5.2) for the Poisson data with the general penalty term $\mathcal{R}$ in Banach spaces has been considered by Werner and Hohage [53, 95]. Analogously, (5.2) for the empirical process data in Banach spaces with general form has also been investigated by Dunker and Hohage [24].

In this chapter we will focus mainly on the convergence rate results of the general theory for (5.2). The general theory for (5.2) has been studied by Pöschl [75] and Flemming [31], we will recall some basic results of regularization theory for (5.2) as presented by Pöschl [75] in Section 5.1. In Section 5.2 we will shortly give an overview for some important results on the convergence rates for inverse problems with Poisson data that have been studied by Werner and Hohage [53, 95] and Antoniadis and Bigot [2]. Furthermore, we will present in Subsection 5.2.2 our main results on the convergence rates for (5.2) by improving the reconstruction error in expectation and the deviation inequalities for stochastic processes we obtained in Chapter 4.

5.1   Regularization properties

For the minimization problem (5.2), we will need the following three common properties:

(a) **Well-posedness:** For any regularization parameter $\alpha > 0$ and observed data $g^{\mathrm{obs}} \in \mathbb{Y}$ there exist at least one minimizer of Tikhonov functional $\mathbf{T}(f)$.

(b) **Stability:** For a fixed parameter $\alpha > 0$, $\widehat{f}_\alpha$ continuously depend on the data $g^{\mathrm{obs}} = G_n$.

(c) **Convergence:** The minimizers of $\mathbf{T}(f)$ converge to an exact solution of equation (5.1) as the noise level and regularization parameter α tend to 0 by choosing α in proper way.

Let us comment on these properties of the minimization problem (5.2). The convergence property (c) is only guarantee that the regularized solution $\widehat{f}_\alpha$ should approximate to the exact solution to (5.1) via (5.2). The items (a) and (b) are not sufficient to derive uniqueness of the minimizers and the problem (5.2) is still ill-posed. Thus, the choice of the regularization parameter α is an important issue to analyze the convergence process.

To derive convergence rates for the method (5.2), the classical definition for the noise level of data fidelity functional $\mathcal{S}(g^{\mathrm{obs}}, g^{\dagger}) \leq \delta$ is not available due to the specific structure of $\mathcal{S}$ in the cases of Poisson and empirical process data and it does not necessarily satisfy a triangle inequality. Therefore, we will apply a effective noise level that we introduced in Chapter 3.

Lemma 5.1. *Let the data fidelity functionals $\mathcal{S}_\zeta(\cdot, g^{\dagger})$ and $\mathcal{S}_\zeta(\cdot, G_n)$ for Poisson and empirical process data given by (2.13) and (2.20) with $\delta > 0$, respectively. Then the noise level*

$$\mathbf{err}(g) = \begin{cases} \int_{\mathbb{M}} \ln\left(\frac{g+\zeta}{g^{\dagger}+\zeta}\right)(\mathrm{d}G_n - g^{\dagger}\mathrm{d}x) & \text{if } g \geq \frac{-\zeta}{2} \text{ a.e.,} \\ \infty & \text{otherwise.} \end{cases} \tag{5.3}$$

Proof. Using the formula of noise level in Definition 3.11, we have

$$\mathbf{err}(g) = \mathcal{S}_\zeta(G_n, g) - \mathcal{S}_\zeta(G_n, g^{\dagger}) - \mathrm{KL}(g + \zeta, g^{\dagger} + \zeta)$$

and by plugging (2.14) and (2.15) into above equation, we immediately derive the assertion (5.3). $\qquad\square$

In general, the modified error term (5.5) can be bounded with high probability by using a concentration inequality if we get $\left\|\ln\left(\frac{g+\zeta}{g^{\dagger}+\zeta}\right)\right\|_\infty < \infty$ under some proper assumption. However, such kind of concentration inequality derived by Reynaud-Bouret [76] cannot be applied for unbounded density functions. For this reason, Werner and Hohage in [95] and Dunker and Hohage [24] bound the log-function in a Sobolev ball. Hohage et al [24, 95] derived the following deviation inequalities for the noise error **err** based on the results of Lemma 2.4 and Lemma 2.9 in Sobolev norms, where they stated their results as follows:

Lemma 5.2 (Deviation inequality in Sobolev norm). *Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded Lipschitz domain and let G_n be either a temporally normalized Poisson process or scaled empirical process with intensity $g^{\dagger} \in L^\infty(\mathbb{M})$ and let $H^s(\mathbb{M})$ denote the L^2-based Sobolev space with index function $s > \frac{d}{2}$ and consider the ball*

$$B_s(R) := \left\{ \mathbf{g} \in H^s(\mathbb{M}) \mid \|\mathbf{g}\|_{H^s(\mathbb{M})} \leq R \right\}.$$

Then there exists a constant $c > 0$ depending on $\mathbb{M}, s$ and $\|g^{\dagger}\|_{L^1}$ such that

$$\mathbf{P}\left(\sup_{\mathbf{g} \in B_s(R)} \left| \int_{\mathbb{M}} \mathbf{g}\left(\mathrm{d}G_n - g^{\dagger}\mathrm{d}x\right) \right| \geq \frac{\eta}{\sqrt{n}}\right) \leq \exp\left(-\frac{\eta}{cR}\right)$$

for all $R \geq 1, t \geq 1$ and $\eta \geq cR$. Furthermore, the following inequality

$$\mathbf{P}\left(\mathbf{err}(B_s(R)) \geq \frac{\eta}{\sqrt{n}}\right) \leq \exp\left(-\frac{\eta}{C_{con}}\right) \tag{5.4}$$

holds true for all $R \geq 1, t \geq 1$ and $\eta \geq C_{con}$, where $C_{con} := 2\max\{\zeta^{-\lfloor s \rfloor - 1}, |\ln(R)|\}Rc$.

Proof. See Corollary 4.2 [95]. □

As we see from Lemma 5.7, the deviation inequality (5.6) cannot yield an optimal bound. With this inequality in Sobolev ball, the authors derived abstract convergence rates in terms of an index function and they orders are suboptimal. In order to improve such kind of rates of convergence, we first need a general convergence result under deterministic setting obtained by Werner and Hohage [53], which is an essential step to derive the convergence rates for the variational regularization of inverse problems with Poisson and empirical process data.

5.2 Convergence rates for stochastic process data

In this section we will start with an overview of the convergence rate results for inverse problems with Poisson data achieved recently by Werner and Hohage [95, 53] as well as convergence results for empirical process data obtained by Dunker and Hohage [24]. Moreover, we also present the convergence rate result of linear inverse problems with Poisson data achieved by Antoniadis and Bigot [2], where they obtained optimal convergence rates for wavelet Galerkin approximation methods. Finally, we further discuss a statistical convergence analysis for variational regularization of mildly ill-posed inverse problems under Assumption 3.18 and improve the convergence rates for nonlinear ill-posed problems with stochastic data including Gaussian white noise, Poisson and empirical process data.

5.2.1 Known convergence results

First of all, we will present some important results for Poisson inverse problems studied by Antoniadis and Bigot [2], where they developed wavelet-Galerkin approximation methods in wavelet-based Besov spaces and obtained optimal convergence rates for linear statistical inverse problems with Poisson data. To derive the optimal rates for Galerkin approximation method, they stated an important assumption.

Assumption 5.3. *Let $\mathbb{M} = (\mathbb{R}/\mathbb{Z})^d$ be d-dimensional measurement manifold. The wavelet functions ψ_λ and $\widetilde{\psi}_\lambda$ are chosen.*

(a) Let $\{\psi_\lambda\}_{|\lambda|=j}$ be a Wavelet basis with $\lambda = (j,k)$ and $\mathbb{X}_j = \operatorname{span}\{\psi_\lambda| \ |\lambda| < j\}$. Then the subspaces $\mathbb{X}_0 \subset \mathbb{X}_1 \subset \cdots \subset L^2(\mathbb{M})$ with 2^{jd} dimensions of $\mathbb{X}_j$ and $d = \dim(\mathbb{M})$. Let $\{\widetilde{\psi}\}_{|\lambda|=j}$ also be an orthonormal basis of the complement $\mathbb{X}_j^c$ of $\mathbb{X}_j$ into $\mathbb{X}_{j+1}$.

(b) For any $P \in L^2(\mathbb{M})$, its wavelet decomposition can be represented by

$$P = \sum_{|\lambda|=0} \langle P, \psi_\lambda \rangle \psi_\lambda + \sum_{j=0}^{\infty} \sum_{|\lambda|=j} \langle P, \widetilde{\psi} \rangle \widetilde{\psi}.$$

For simplicity, we may write $\{\psi_\lambda\}_{|\lambda|=-1}$ for $\{\widetilde{\psi}_\lambda\}_{|\lambda|=0}$ and consider j is starting from -1.

(c) Assume that $\|\widetilde{\psi}\|_{L^\infty(\mathbb{M})} = 2^{|\lambda|d/2} \|\widetilde{\psi}\|_{L^\infty(\mathbb{M})}$.

(d) For all $j \in \mathbb{N}$ there exists A_j such that for all $x \in \mathbb{X}_j$

$$\|x\|_{L^\infty(\mathbb{M})} \leq A_j \|x\|_{L^2(\mathbb{M})}.$$

Under Assumption 5.3, the wavelets provide an unconditional basis for Besov spaces, the wavelet based Besov space norm is given by (4.3). If Assumption 5.3 holds true, then the linear Galerkin approximation estimator of $f^\dagger$ is given by

$$\langle Tf_j^G, \widetilde{\psi} \rangle = \langle g^{\mathrm{obs}}, \widetilde{\psi} \rangle \quad \text{for all } |\lambda| < j,$$

where the function $f_j \in \mathbb{X}_j$ satisfying this equation with $g^{\mathrm{obs}} = g^\dagger$ is called the **Galerkin approximation** of $f^\dagger$. Now we define the Galerkin wavelets $g_\lambda^j \in \mathbb{X}_j$ by

$$\langle Tg_\lambda^j, x \rangle = \langle \psi_\lambda, x \rangle$$

for all $x \in \mathbb{X}_j$. Moreover, if we define a standard linear projection estimator $f_j \in \mathbb{X}_j$ by

$$\langle f_j, \widetilde{\psi}_\lambda \rangle = \int_{\mathbb{M}} g_\lambda^j \mathrm{d}G_n \tag{5.5}$$

for all $|\lambda| < j$, then f_j is the Galerkin approximation of $f^\dagger$. However, the estimator (5.5) is not necessarily non-negative. Therefore, the Galerkin information projection of $f^\dagger$ onto the exponential family $\mathbb{E}_j$ at scale j can be defined as a function space of the form

$$\mathbb{E}_j = \left\{ f_{j,\theta} = \exp\left(\sum_{|\lambda|<j} \theta_\lambda \psi_\lambda\right), \theta = (\theta_\lambda)_{|\lambda|<j} \in \mathbb{R}^{2^{jd}} \right\}. \tag{5.6}$$

With this approximation we can guarantee that $f_{j,\theta}$ is positive. Csiszár [21] show that

$$\langle f_{j,\theta}, \widetilde{\psi}_\lambda \rangle = \langle f^\dagger, \widetilde{\psi}_\lambda \rangle$$

for any $|\lambda| < j$. To estimate the intensity function $f^\dagger$, Antoniadis and Bigot [2] proposed to define the estimator $f_{j,\hat{\theta}} \in \mathbb{E}_j$ such that

$$\langle f_{j,\hat{\theta}}, \psi_\lambda \rangle = \int_{\mathbb{M}} g_\lambda^j \mathrm{d}G_n \tag{5.7}$$

for all $|\lambda| < j$, where $f_{j,\hat{\theta}}$ is called *Galerkin information projection estimate* of $f^\dagger$ if there exists a solution to this problem. Note that the estimator $f_{j,\hat{\theta}}$ depends linearly on G_n while the corresponding parameter $\hat{\theta} \in \mathbb{R}^{2^{jd}}$ doesn't dependen on G_n due to the set (5.6). The following results for the asymptotic behavior of the estimator (5.7) were stated by Antoniadis and Bigot [2]:

Theorem 5.4 (Optimal rate for linear estimator). *Let* $\mathbb{M} = (\mathbb{R}/\mathbb{Z})^d$ *be d-dimensional measurement manifold and the linear operator* $T : L^2(\mathbb{M}) \to L^2(\mathbb{M})$ *is self-adjoint, positive definite and an isomorphism between* $H^s(\mathbb{M})$ *and* $H^{s+\nu}(\mathbb{M})$ *for any* $\tau \geq 0$ *for some* $\nu > 0$ *and* $f^\dagger = e^{h^\dagger}$ *with some* $h^\dagger \in H^s(\mathbb{M}) = B_{2,2}^s(\mathbb{M})$ *where* $s > d/2$. *Assume that* ψ *is compactly supported and* $\psi \in H^{s+\nu+d/2}$ *with* $s > \nu - d/2$ *and has vanishing moments with* $r > s + d/2$. *Then for a priori parameter choice*

$$2^{-j(n)} = \left(\frac{1}{n}\right)^{\frac{1}{2\nu}},$$

with probability tending to 1 as $n \to \infty$ *the Galerkin information projection exist and satisfies*

$$\mathbf{E}\left[KL\left(f^\dagger, f_{j,\hat{\theta}}\right)\right] = O\left(\left(\frac{1}{n}\right)^{\frac{2s}{2s+2\nu+d}}\right), \tag{5.8}$$

as $n \to \infty$.

Proof. See Theorem 4.2 [2]. ☐

Note that the convergence rate result (5.8) is concerning projection with respect to the Kullback-Leibler divergence in pre-image space $\mathbb{X}$, whereas Werner and Hohage [95] considered that the Kullback-Leibler divergence appears in connection with the log-likelihood data fidelity functional of the Poisson distribution in $\mathbb{Y}$ space.

Moreover, the linear estimators $f_{j,\hat{\theta}}$ defined in Theorem 5.4 unfortunately do not yield the optimal rates when the exact solution $f^\dagger \in B_{p,p}^s(\mathbb{M})$ with $p \in [1,2)$. It is also true for inverse problems with additive Gaussian white noise, see [23]. To overcome such difficulties, Antoniadis and Bigot [2] considered nonlinear estimators. Recall that the coefficients defining the Galerkin information projection estimate of f at scale j are given by

$$\hat{f}_j^{nl} = \sum_{|\lambda| < j} U_\epsilon \left(\int_{\mathbb{M}} g_\lambda^j dG_n\right) \psi_\lambda, \tag{5.9}$$

where the estimator $\hat{f}_j^{nl}$ is not necessarily non-negative for the linear case. Thus the **Galerkin information projection** onto the set $\mathbb{E}_j$ can be defined by

$$\left\langle f_{j,\hat{\theta}}^{\text{nl}}, \psi_\lambda \right\rangle = \left(C_\lambda^j\right)^U \left(U_\epsilon \left(\int_{\mathbb{M}} g_\lambda^j dG_n\right)\right)_{|k| < j}$$

for all $|\lambda| < j$, where $\left(C_\lambda^j\right)^U$ denotes the linear transformation mapping the coefficients with respect to $\{\widetilde{\psi}_\lambda\}$ to the λ-th coefficient with respect to $\{g_\lambda^j\}$.

Theorem 5.5 (Optimal rates for nonlinear estimator). *Supoose that* $\mathbb{M} = (\mathbb{R}/\mathbb{Z})^d$ *be d-dimensional measurement manifold and the linear operator* T *is self-adjoint, positive definite and an isomorphism between* $L^2(\mathbb{M})$ *and* $H^\nu(\mathbb{M})$ *and maps* $B_{p,p}^\tau(\mathbb{M})$ *bounded into* $B_{p,p}^{s+\nu}(\mathbb{M})$ *for any* $s \geq 0$ *for some* $\nu > 0$ *and* $f^\dagger = e^{h^\dagger}$ *with some* $h^\dagger \in B_{p,p}^s(\mathbb{M})$ *where* $s > 0$ *and* $1/p = 1/2 + s(2\nu + d)$. *Moreover, suppose that Assumption 5.3 is satisfied and* ψ *is compactly supported and* $\psi \in H^{\tau+\nu+d/2}$ *with* $s > \nu - d/2$ *and has vanishing moments with* $r > s + d/2$. *Then for a priori parameter choice*

$$2^{-j(n)} \leq \left(\frac{1}{n}\right)^{\frac{1}{2s+2\nu+d}}, \quad \epsilon = \frac{\sqrt{|\ln(n)|}}{\sqrt{n}} 2^{\nu|\lambda|}$$

with probability tending to 1 as $n \to \infty$ the Galerkin information projection exist and satisfies

$$\mathbf{E}\left[KL\left(f^\dagger, f_{j,\hat{\theta}}^{nl}\right)\right] = O\left(\left(\frac{\sqrt{|\ln(n)|}}{n}\right)^{\frac{2s}{2s+2\nu+d}}\right), \tag{5.10}$$

as $n \to \infty$.

Proof. See Theorem 5.1 [2]. □

Although the rates of convergence (5.8) and (5.10) for inverse problems with Poisson data shown in Theorem 5.4 and Theorem 5.5 are known to be order optimal, their approach was restricted to inverse problems in identical Borel subsets of $\mathbb{R}^d$ as well as the operator F is linear and self-adjoint positive definite. However, the optimal rates of convergence for such inverse problems with Poisson data in the case of non-linear operators remain open so far.

In recent years, Werner and Hohage [95, 53] intensively studied convergence analysis for Tikhonov-type regularization of non-linear ill-posed inverse problems with Poisson data under some sort of source conditions. To measure the smoothness of the unknown solution, we intend to use source conditions in the form of variational inequalities and the penalty functional $\mathcal{R}$ given by the quadratic Hilbert space norms, which has been introduced in Assumption 3.18 with the general form in Chapter 3.

The following theorem can be seen as a particular case of Theorem 4.11 derived by Hohage and Werner [53], where we use the error distance between the estimator $\widehat{f}_\alpha$ and the exact solution $f^\dagger$ in the quadratic Hilbert space norms instead of general Bregman distance due to the specific type of penalty functional $\mathcal{R}$.

Theorem 5.6 (A priori rates of convergence). *Suppose that Assumption 3.18 holds true. If $F(f) \geq 0$ for all $f \in \mathcal{B} \subset \mathbb{X}$ and if the Tikhonov functional (5.2) has a global minimizer $\widehat{f}_\alpha$, then we obtain*

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq \frac{2\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + 2\Psi(\alpha) \tag{5.11}$$

and

$$\mathcal{T}_\zeta(F(\widehat{f}_\alpha), g^\dagger) \leq \mathbf{err}(F(\widehat{f}_\alpha)) + 2\Psi(2\alpha) \tag{5.12}$$

for all $\alpha > 0$. Moreover $F(\mathcal{B})$ 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

$$-\frac{1}{\alpha} \in \partial(-\Phi)\left(\frac{1}{\sqrt{n}}\right) \tag{5.13}$$

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

$$\mathbf{E}\left[\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2\right] = O\left(\Phi\left(\frac{1}{\sqrt{n}}\right)\right) \tag{5.14}$$

as $n \to \infty$.

Proof. See Theorem 4.11 [53]. □

Convergence rate results of the risk for penalized maximum-likelihood estimators also derived by Dunker and Hohage [24]. They studied the convergence analysis for variational regularization of inverse problems with empirical process data, see Theorem 6 and 7 in [24].

Unfortunately, these known results for variational regularization methods of non-linear ill-posed problems cannot yield the optimal rates like the results in (5.8). Due to the abstract structure of variational source conditions, the convergence result in Theorem 5.6 that the smoothness of the true solution $f^\dagger$ is characterized by an index function Φ, in which case they could derive an abstract convergence rate in terms of the index function Φ and smoothing properties of the operator F. Weidling et al [93] recently developed a method for the verification of variational source conditions in Hilbert spaces under standard smoothness conditions, see Theorem 4.1 in [93]. The following lemma will precisely describe the variational source condition with specific powers.

Lemma 5.7. *Let Assumption 5.11 is satisfied and $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau} \leq \varrho$. Then there exists a constant $c > 0$ such that the solution $f^\dagger$ satisfies the variational source conditions (3.35) with*

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

Moreover, Ψ is given by

$$\Psi(\alpha) = c\varrho^2 \alpha^{\frac{\tau}{u}}. \tag{5.15}$$

Proof. The proof of the first argument follows from Theorem 4.1 in [93]. Now we need to calculate the Fenchel conjugate of Φ. By Fenchel formula, we have

$$\Psi(\gamma) = \sup_{\alpha > 0} \left[c\varrho^{\frac{2u}{u+\tau}} \alpha^{\frac{\tau}{\tau+u}} - \frac{\alpha}{\gamma} \right]. \tag{5.16}$$

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

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

Then by inserting this into (5.16) we obtain (5.15). $\square$

Remark 5.8. *Under the assumptions of Lemma 5.7 and with parameter choice*

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

we can write the convergence rate (5.16) more explicit form as follows

$$\mathbf{E}\left[\left\| \hat{f}_\alpha - f^\dagger \right\|_{\mathbb{X}}^2 \right] = O\left(\varrho^{\frac{4u}{2\tau+2u}} \left(\frac{1}{n} \right)^{\frac{\tau}{2\tau+2u}} \right), \tag{5.17}$$

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 (5.17) is not satisfied the order optimality of convergence rates for mildly ill-posed inverse problems.

Note that the choice rule (5.13) is not available in practice due to the fact that it requires a *priori* knowledge of the function Φ characterizing abstract smoothness of the unknown solution $f^\dagger$. Therefore, Werner and Hohage [95] applied a Lepskii-type parameter choice rule to select a regularization parameter $\alpha_{j_{bal}}$ among a number of finite regularization parameters $\alpha_1, \cdots, \alpha_m$, which was first time suggested by Lepskii [63] under a statistical setting. However, the well-known discrepancy principle for inverse problems with Poisson and empirical process data is not applicable due

to the continuous setup. The knowledge of the index function Φ no longer required if the balancing principle is used. The main purpose of the Lepskii balancing principle is to find a regularization parameter $\alpha_{j_{bal}}$ among a number of finite regularization parameters $\alpha_1, \cdots, \alpha_m$, this yields almost optimal convergence rates.

In this setting we only consider the convergence results in the Hilbert spaces with quadratic penalty term $\mathcal{R}(f) = \frac{1}{2}\|f\|_{\mathbb{X}}^2$ such that the inequality in Lemma 3.17 hold true with $q = 2$ and $C_{bd} = 1$. The following theorem is a particular case of Theorem 5.1 in [95].

Theorem 5.9 (A posteriori rates of convergence). *Suppose that Assumption 3.18 and (3.34) hold true with $\beta \in [\frac{1}{2}, \infty)$ and $\Phi^{1+\varepsilon}$ is concave for $\varepsilon > 0$. If $F(f) \geq 0$ for all $f \in \mathcal{B} \subset \mathbb{X}$ and $F(\mathcal{B})$ is bounded with respect to $B_{2,1}^u(\mathbb{M})$-norm topology $\tau_{\mathbb{Y}}$ for $u > d/2$ and if the Tikhonov functional (5.2) has a global minimizer $\widehat{f}_\alpha$ and $\delta > 0, \tau \geq \frac{1}{2}C_{\text{con}}$ with C_{con} as in Lemma 5.2. Consider the regularization parameter α_j by*

$$\alpha_j = \frac{\tau \ln(n)}{\sqrt{n}} e^{2j-1}, \quad j \in \mathbb{N}$$

with some number $e > 1$. Then with $m = \min\{j \in \mathbb{N} : \alpha_j \geq 1\}$ the parameter choice

$$j_{\text{bal}} := \min \left\{ j \in \{1, \cdots, m\} \,\middle|\, \left\|\widehat{f}_{\alpha_j} - \widehat{f}_{\alpha_j}\right\|_{\mathbb{X}}^2 \leq 8e^{1-j} \quad \text{for all } i < j \right\},$$

we obtain

$$\mathbf{E}\left[\left\|\widehat{f}_{\alpha_{bal}} - f^\dagger\right\|_{\mathbb{X}}^2\right] = O\left(\Phi\left(\frac{\ln(n)}{\sqrt{n}}\right)\right) \quad \text{as } n \to \infty. \tag{5.18}$$

Proof. See Theorem 5.1 [95]. $\qquad\square$

Note that the regularization parameter α_{bal} selected by the Lepskii-type parameter choice rule with $\alpha_{bal} = \alpha_{j_{bal}}$. It can be seen from (5.10) and (5.18) that in some cases the loss of such a logarithmic factor is not evitable for adaptation to unknown smoothness.

5.2.2 Improved convergence rates for stochastic data

As we have seen in section 5.2.1, the convergence rate results of Theorem 5.6 and 5.9 for the variational regularization method of non-linear mildly ill-posed inverse problems with Poisson and empirical process data are known to be not optimal. Therefore, in order to improve such suboptimal convergence rates, we further investigate better reconstruction error bounds in expectation for the variational regularization method with the result of deviation inequalities in Chapter 4 and deterministic convergence results of (5.11) in Theorem 5.6. Before we state our main results on convergence rates of inverse problem with Poisson and empirical process data, we need the following small lemma, which will be used to bound the error term.

Lemma 5.10. *Suppose that $b > a > 0$ and $R > 0$ and let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $u \in B_{2,2}^s(\mathbb{M})$ with $s \geq \frac{d}{2}$ and $u \geq a > 0$, then $\ln(u) \in B_{2,2}^s(\mathbb{M})$ and $u^{-1} \in B_{2,2}^s(\mathbb{M})$. Moreover, there exists a constant $\widetilde{C}$ such that*

$$\|\ln(u)\|_{B_{2,2}^s(\mathbb{M})} \leq \widetilde{C}\|u - 1\|_{B_{2,2}^s(\mathbb{M})} \tag{5.19}$$

holds true for all $u \in B_{2,2}^s(\mathbb{M})$ such that $u(x) \in [a,b]$ for almost x and $\|u-1\|_{B_{2,2}^s(\mathbb{M})} \leq R$. Moreover,

$$\left\| \frac{1}{u} \right\|_{B_{2,2}^s(\mathbb{M})} \leq \tilde{C} \left(\frac{1}{a} \|\mathbf{1}\|_{B_{2,1}^s(\mathbb{M})} + \|u-a\|_{B_{2,2}^s(\mathbb{M})} \right) \tag{5.20}$$

for all $u \geq a$ with $\|u\|_{B_{2,2}^s(\mathbb{M})} \leq R$, where $\mathbf{1}$ denotes the constant function with value 1.

Proof. Set $v := u - 1$ and define a composition operator $T_H(v) = H \circ v$ on Besov spaces $B_{2,2}^s(\mathbb{M})$ associated with a function $H : \mathbb{R} \to \mathbb{R}$ given by $H(x) = \ln(x+1)$ for $x \in [a+1, b+1]$. Then (5.19) is equivalent to

$$\|H \circ v\|_{B_{2,2}^s(\mathbb{M})} \leq \widetilde{C} \|v\|_{B_{2,2}^s(\mathbb{M})} \tag{5.21}$$

for all $v \in B_{2,2}^s(\mathbb{M})$ satisfying $v + 1 \in [a,b]$ and $\|v\|_{B_{2,2}^s(\mathbb{M})} < R$. The function H can be extended from the interval $[a-1, b-1]$ 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

$$\|H \circ v\|_{B_{2,2}^s(\mathbb{R}^d)} \leq c \|H\|_{C^{m+1}} \left(\|v\|_{B_{2,2}^s(\mathbb{R}^d)} + \|v\|_{B_{2,2}^s(\mathbb{R}^d)}^{\tilde{s}} \right) \tag{5.22}$$

for all $v \in B_{2,2}^s(\mathbb{R}^d)$ with $\tilde{s} = \max(1,s)$ and for $c > 0$. Then the inequality (5.22) shows (5.21) with constant $\tilde{C} :=: c \|H\|_{C^{m+1}}(1 + R^{\tilde{s}-1})$ and $\mathbb{M} = \mathbb{R}^d$. In the case $\mathbb{M} \neq \mathbb{R}^d$, we use that $B_{2,2}^s(\mathbb{M}) = \{f|_{\mathbb{M}} : f \in B_{2,2}^s(\mathbb{R}^d)\}$ with norm by the infimum over $\|f\|_{B_{2,2}^s(\mathbb{R}^d)}$. As $(H \circ v)|_{\mathbb{M}} = H \circ v|_{\mathbb{M}}$, then we obtain the assertion for general $\mathbb{M}$.

The second statement is proven analogously by using the function $H(x) = \frac{1}{x+a} - \frac{1}{a}$ for $a > 0$. Hence, by applying Theorem A in Adams and Frazier [1] once again, we derive (5.20). $\qquad \square$

Assumption 5.11 (Smoothing and Lipschitz properties). *Let $\mathbb{M} \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Let $F : \mathcal{B} \subset \mathbb{X} \to B_{2,2}^u(\mathbb{M})$ 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 $L > 0$.

It is necessary to emphasize that the smoothness assumption implies the boundedness of F from $\mathbb{X}$ to the Besov space $B_{2,2}^s(\mathbb{M})$ if $F : \mathbb{X} \to B_{2,2}^s(\mathbb{M})$ is linear. Before we present and prove the main results, it is necessary to recall the variational inequalities for the case of data fidelity functional $\mathcal{S}$ being the Kullback-Leibler divergence $KL(g, g^\dagger)$ with shifted version, which has been proven by Borwein and Lewis [13].

Lemma 5.12 (Lower bound for $\mathcal{T}_\zeta$). *Let $\zeta \geq 0$ and $g, g^\dagger \in L^\infty(\mathbb{M})$ and $\mathcal{T}_\zeta(g, g^\dagger)$ be defined by (2.14) and (2.21). Then the following estimates hold true for all $g \geq 0$ such that*

$$\|g - g^\dagger\|_{L^2(\mathbb{M})}^2 \leq \left(\frac{2}{3} \|g^\dagger\|_{L^\infty(\mathbb{M})} + \frac{4}{3} \|g\|_{L^\infty(\mathbb{M})} + 2\zeta \right) \mathcal{T}_\zeta(g, g^\dagger). \tag{5.23}$$

Proof. See Lemma 4.5 [94]. $\qquad \square$

Lemma 5.13. *Suppose that Assumption 5.11 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 (5.2), then there exists a*

constant $\widetilde{C}_{con} > 0$ such that

$$\left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,1}^s(\mathbb{M})} \leq \widetilde{C}_{con} \| \widehat{f}_\alpha - f^\dagger \|_{\mathbb{X}}^{\frac{s}{2u}} \mathcal{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}} \tag{5.24}$$

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

Proof. To bound the logarithmic function in (5.24), we will first use the interpolation inequality in Besov spaces with the domain $\mathbb{M}$. By Theorem 1.33 in Triebel [88], we have

$$\left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,1}^s(\mathbb{M})} \leq C_1 \left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,2}^u(\mathbb{M})}^{\frac{u}{s}}$$

$$\times \left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,2}^0(\mathbb{M})}^{1-\frac{u}{s}} \tag{5.25}$$

for some constant $C_1 > 0$ and $u > s > 0$. Note that $B_{2,2}^u(\mathbb{M}) = H_2^u(\mathbb{M})$ and $B_{2,2}^0(\mathbb{M}) = L^2(\mathbb{M})$ for the smoothness $u > s > 0$. As a consequence of Lemma 5.10 and by Theorem 2.8.3 in Triebel [89] for the bounded Lipschitz domain $\mathbb{M} \subset \mathbb{R}^d$, the first factor on the right hand side of (5.25) can be bounded by

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

$$\leq C_2 \left(\frac{2(\|g^\dagger\|_{L^\infty} + \delta)}{\delta} \right)^m \left\| \frac{1}{g^\dagger + \delta} \right\|_{B_{2,2}^u(\mathbb{M})}$$

$$\times \left\| F(\widehat{f}_\alpha) - g^\dagger \right\|_{B_{2,2}^u(\mathbb{M})} \tag{5.26}$$

with some constant $C_2 > 0$, where we used the inequality

$$\frac{g + \delta}{g^\dagger + \delta} \geq \frac{\delta}{2(\|g^\dagger\|_{L^\infty} + \delta)} \quad \text{for small} \quad \delta \leq 1.$$

The second factor on the right hand side of (5.26) is bounded based on the result (5.20) of Lemma 5.10. Moreover, by Assumption 5.11, we have

$$\|F(\widehat{f}_\alpha) - g^\dagger\|_{B_{2,2}^u(\mathbb{M})} \leq L \|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}} \tag{5.27}$$

for Lipschitz constant $L > 0$. From the inequality (5.26) and (5.27) we have

$$\left\| \ln \left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,2}^u(\mathbb{M})} \leq C_{\delta, g^\dagger, L} \left\| \widehat{f}_\alpha - f^\dagger \right\|_{\mathbb{X}} \tag{5.28}$$

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 (5.25). Similarly, with the help of Lemma 5.10 we have

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

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

$$
= C_{\delta,g^\dagger} \left\| F(\widehat{f}_\alpha) - g^\dagger \right\|_{L^2(\mathbb{M})} \tag{5.29}
$$

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 5.12 for the norm $\|F(\widehat{f}_\alpha) - g^\dagger\|_{L^2(\mathbb{M})}$, we obtain

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

$$
\times \mathcal{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}, \tag{5.30}
$$

where $\|F(f)\|_{L^\infty(\mathbb{M})}$ is bounded due to the fact that $\sup_{f \in \mathcal{B}} \|F(f)\|_{B_{2,2}^s(\mathbb{M})}$ 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 (5.30) is bounded by

$$
\|F(\widehat{f}_\alpha) - g^\dagger\|_{L^2(\mathbb{M})} \leq C_3 \mathcal{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}. \tag{5.31}
$$

By plugging (5.31) into (5.29) we have

$$
\left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \delta}{g^\dagger + \delta} \right) \right\|_{B_{2,2}^0(\mathbb{M})} \leq C_{\delta,g^\dagger,C_3} \mathcal{T}_\delta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{1}{2}}. \tag{5.32}
$$

Finally, by inserting (5.28) and (5.32) into (5.25) we obtain

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

Setting $\widetilde{C}_{con} := C_1 C_2 C_{\delta,g^\dagger,L} C_{\delta,g^\dagger,C_3} > 0$, which yields the assertion. [1]

$\square$

Theorem 5.14. *Let* $F : \mathcal{B} \subset \mathbb{X} \to L^\infty(\mathbb{M})$ *be a continuous operator and and* $F(f) \geq 0$ *for all* $f \in \mathcal{B} \subset \mathbb{X}$ *and* $F(\mathcal{B})$ *is bounded in* $B_{2,2}^s(\mathbb{M})$ *and* $\sup_{f \in \mathcal{B}} \|F(f)\|_{B_{2,2}^s(\mathbb{M})} < \infty$ *with some* $s \in [d/2, u)$. *Suppose that Assumptions 3.18 and 5.11 hold true. Let* $\widehat{f}_\alpha$ *be a minimizer of*

---

[1]This lemma was improved in the revision of this thesis based on ideas by Benjamin Sprung published in his PhD thesis. By first interpolating and then using the composition lemma 5.10 rather than the other way around we are now able to avoid unnecessary restrictions on the smoothness index s.

(1.3). Then

$$\mathbf{err}(F(\widehat{f}_\alpha)) \le 2^{\frac{s}{u}}\widetilde{C}_s^2 \alpha^{-\frac{s}{u}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2 + \alpha\Psi(2\alpha) \tag{5.33}$$

and

$$\left\| \widehat{f}_\alpha - f^\dagger \right\|_{\mathbb{X}}^2 \le 2^{\frac{s}{u}}\widetilde{C}_s^2 \alpha^{-1-\frac{s}{u}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2 + \Psi(2\alpha) \tag{5.34}$$

for all $\alpha > 0$ and some constant $\widetilde{C}_s > 0$.

Proof. By Hölder's inequality and duality Theorem 4.6 in Chapter 4 and Lemma 5.13, the error term (5.5) can be bounded by

$$\begin{aligned}
\mathbf{err}(F(\widehat{f}_\alpha)) &= \left| \int_{\mathbb{M}} \ln\left(\frac{F(\widehat{f}_\alpha) + \zeta}{g^\dagger + \zeta} \right) \left(dG_n - g^\dagger dx \right) \right| \\
&\le \| G_n - g^\dagger \|_{B_{2,\infty}^{-s}(\mathbb{M})} \left\| \ln\left(\frac{F(\widehat{f}_\alpha) + \zeta}{g^\dagger + \zeta} \right) \right\|_{B_{2,1}^s(\mathbb{M})} \\
&\le \widetilde{C}_s \| G_n - g^\dagger \|_{B_{2,\infty}^{-s}(\mathbb{M})} \| \widehat{f}_\alpha - f^\dagger \|_{\mathbb{X}}^{\frac{s}{u}} \mathcal{T}_\zeta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}}, \tag{5.35}
\end{aligned}$$

where

$$\| G_n - g^\dagger \|_{B_{2,\infty}^{-s}(\mathbb{M})} := \sup_{\|\mathbf{F}\|_{B_{2,1}^s(\mathbb{M})} \le 1} \left| \int_{\mathbb{M}} \mathbf{F}(x)(dG_n - g^\dagger dx) \right|$$

as $B_{2,\infty}^{-s}(\mathbb{M})$ is the dual space of $B_{2,1}^s(\mathbb{M})$. By the monotoncity of Ψ and the error bound (5.11), we have

$$\frac{1}{2}\|\widehat{f}_\alpha - f^\dagger\|_{\mathbb{X}}^2 \le \frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + \Psi(\alpha) \le \frac{1}{\alpha}\left(\mathbf{err}(F(\widehat{f}_\alpha)) + \alpha\Psi(2\alpha) \right). \tag{5.36}$$

By plugging the deterministic error bound (5.11) in Theorem 5.6 and (5.36) into (5.35), using the inequality $(x+y)^{\frac{1}{p}} \le x^{\frac{1}{p}} + y^{\frac{1}{p}}$ and Young's inequality $xy \le \frac{1}{p}x^p + \frac{1}{q}y^q$ for all $x, y > 0$ with $p, q > 0$ satisfies $\frac{1}{p} + \frac{1}{q} = 1$, then the error (5.35) is further bounded by

$$\begin{aligned}
\mathbf{err}(F(\widehat{f}_\alpha)) &\le \widetilde{C}_s \| \widehat{f}_\alpha - f^\dagger \|_{\mathbb{X}}^{\frac{s}{u}} \mathcal{T}_\zeta(F(\widehat{f}_\alpha), g^\dagger)^{\frac{u-s}{2u}} \| G_n - g^\dagger \|_{B_{2,\infty}^{-s}} \\
&\le 2^{\frac{s}{2u}}\widetilde{C}_s \alpha^{\frac{u-s}{2u}} \left(\frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} + \Psi(2\alpha) \right)^{\frac{1}{2}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}} \\
&\le \widetilde{C}_s 2^{\frac{s}{2u}} \alpha^{\frac{u-s}{2u}} \left(\frac{\mathbf{err}(F(\widehat{f}_\alpha))}{\alpha} \right)^{\frac{1}{2}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}} \\
&\quad + \widetilde{C}_s 2^{\frac{s}{2u}} \alpha^{\frac{u-s}{2u}} \Psi(2\alpha)^{\frac{1}{2}} \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}} \\
&\le \frac{1}{2}\mathbf{err}(F(\widehat{f}_\alpha)) + \frac{\alpha}{2}\Psi(2\alpha) + \frac{1}{2} \cdot 2^{\frac{s}{u}}\alpha^{\frac{u-s}{u}} \left(\frac{\widetilde{C}_s}{\sqrt{\alpha}} \right)^2 \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2 \\
&\quad + \frac{1}{2} \cdot 2^{\frac{s}{u}}\alpha^{\frac{u-s}{u}} \left(\frac{\widetilde{C}_s}{\sqrt{\alpha}} \right)^2 \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}}^2, \tag{5.37}
\end{aligned}$$

by subtracting $\frac{1}{2}\mathbf{err}(F(\widehat{f}_\alpha))$ on the both side of the inequality (5.37), we obtain

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

Finally, by inserting this inequality into (5.33), we immediately obtain (5.34). $\qquad\square$

Remark 5.15. *Obviously, it is possible to extend the the error bound (5.33) and (5.34) in Banach space setting. Here we only consider the penalty functionals with quadratic form, i.e., $\mathcal{R}(f) = \frac{1}{2}\|f\|_\mathbb{X}^2$. Similar results also hold true for Banach spaces with $\mathcal{R}(f) = \frac{1}{q}\|f\|_\mathbb{X}^q$ for $q \geq 2$.*

Theorem 5.16 (**Error bounds for Gaussian white noise**). *Assume that $F : \mathcal{B} \subset \mathbb{X} \to L^\infty(\mathbb{M})$ be a continuous operator and Assumptions 3.18 and 5.11 hold true with $u > d/2$. Let $\widehat{f}_\alpha$ be a minimizer of (5.2). Then*

$$\mathbf{err}(F(\widehat{f}_\alpha)) \leq 2^{\frac{s}{u}}\widetilde{C}_s^2\alpha^{-\frac{s}{u}}\varepsilon\|W\|_{B_{2,\infty}^{-s}}^2 + \alpha\Psi(2\alpha) \tag{5.38}$$

and

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_\mathbb{X}^2 \leq 2^{\frac{s}{u}}\widetilde{C}_s^2\alpha^{-1-\frac{s}{u}}\varepsilon\|W\|_{B_{2,\infty}^{-s}}^2 + \Psi(2\alpha) \tag{5.39}$$

for all $\alpha > 0$ and for $s \geq \frac{d}{2}$ and some constant $\widetilde{C}_s > 0$.

Proof. For the Gaussian white noise model, the error functional (3.30) with $g = F(\widehat{f}_\alpha)$ can be bounded as in Besov norm

$$\mathbf{err}(F\widehat{f}_\alpha) = \varepsilon\langle W, F(\widehat{f}_\alpha) - g^\dagger\rangle \leq \varepsilon\|W\|_{B_{2,\infty}^{-\frac{d}{2}}(\mathbb{M})}\|F(\widehat{F}_\alpha) - g^\dagger\|_{B_{2,1}^{\frac{d}{2}}(\mathbb{M})}.$$

The remainder of the proof goes along the same line as (5.35) and (5.36). $\qquad\square$

Now we are in the position to prove the improved convergence rates for variational regularization of nonlinear inverse problems with Poisson and empirical process data simultaneously.

Theorem 5.17 (**Convergence rates for Poisson and empirical process data**). *Let $F : \mathcal{B} \subset L^2(\mathbb{M}) \to B_{2,1}^u(\mathbb{M})$ be a continuous operator and $F(f) \geq 0$ for all $f \in \mathcal{B}$. Assume that $F(\mathcal{B})$ is bounded in $B_{2,1}^s(\mathbb{M})$ and $\sup_{f\in\mathcal{B}}\|F(f)\|_{B_{2,1}^s(\mathbb{M})} < \infty$ with some $s \in [d/2, u)$. Suppose that Assumptions 3.18 and 5.11 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 $\|f^\dagger\|_{B_{2,\tau}^\tau(\mathbb{M})} \leq \varrho$ and for the parameter choice rule*

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

the error bound

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_\mathbb{X}^2 \leq C_1\varrho^{\frac{4\tau+2\tilde{d}}{2\tau+2a+\tilde{d}}}n^{-\frac{2\tau}{2\tau+2u+\tilde{d}}}\left(1 + n\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2\right) \tag{5.40}$$

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

$$\mathbf{E}\left[\left\|\widehat{f}_{\alpha^*} - f^\dagger\right\|_\mathbb{X}^2\right]^{1/2} \leq C_2\varrho^{\frac{2\tau+\tilde{d}}{2\tau+2u+\tilde{d}}}n^{-\frac{\tau}{2\tau+2u+\tilde{d}}} \tag{5.41}$$

as $n \to \infty$. Moreover, if Conjecture 4.9 holds true with $\tilde{d} = d$, then we obtain the following convergence rate

$$\mathbf{E}\left[\left\|\widehat{f}_{\alpha^*} - f^\dagger\right\|_{\mathbb{X}}^2\right]^{1/2} \leq C_3(\ln n)^p \varrho^{\frac{2\tau+2d}{2\tau+2u+d}} n^{-\frac{\tau}{2\tau+2u+d}} \tag{5.42}$$

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

Proof. By combining the results of Lemma 5.7 and Theorem 5.16, we have

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

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

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq C_1 \varrho^{\frac{4\tau+2\tilde{d}}{2\tau+2a+d}} n^{-\frac{2\tau}{2\tau+2u+d}} \left(1 + n\|G_n - g^\dagger\|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})}^2\right) \tag{5.43}$$

for all $n \in \mathbb{N}$. Finally, by taking expectation on (5.43) and Theorem 4.11 and 4.12 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 4.9 holds true with $p = 0$ which implies (5.41) and (5.42) for $p = 0$. If $p > 0$, then

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

and we derive (5.42) for $p > 0$, which complete the proof of the theorem. $\qquad\square$

Now let us discuss the main achievements on the convergence results of Theorem 5.4 and Theorem 5.6 with convergence rates of Theorem 5.17. On the one hand, Theorem 5.17 show that the convergence rates for variational regularization methods of nonlinear inverse problems with Poisson and empirical process data are order optimal under the conjecture 4.9 . We derive the convergence rates for variational regularization methods with the explicit exponent $\frac{\tau}{2\tau+2u+d}$, which is known to be order optimal in the sense of minimax sense and the convergence results are not affected by the nonlinearity of the forward operator. However, Antoniadis and Bigot [2] are restricted to linear operators in certain function spaces and the convergence results for the wavelet Galerkin approximation methods only applied linear operators. Theorem 5.17 can be applied for nonlinear ill-posed inverse problems with Poisson and empirical process data.

As we have seen in Chapter 3, we already discussed optimal convergence rates for Tikhonov regularization of linear ill-posed inverse problems with additive random noise based on the spectral theory. Moreover, Burger et al [17] obtained nearly optimal convergence rate for variational regularization of linear inverse problems with additive Gaussian white noise. We are able to achieve the optimal convergence rates for variational regularization of mildly ill-posed nonlinear inverse problems with Gaussian white noise. It is worth emphasizing that the optimal convergence rates for Gaussian white noise are very easily obtained due to the deviation inequality (4.11) for all $p \in [1, \infty)$. Therefore, we will summarize the optimal convergence results for Gaussian white noise as follows:

Corollary 5.18 (Order optimal of convergence rates for Gaussian white noise). *Let $F : \mathcal{B} \subset L^2(\mathbb{M}) \to B_{2,1}^u(\mathbb{M})$ be a continuous operator. Suppose that Assumptions 3.18 and 5.11 hold true with the norm topology for index $u > \frac{d}{2} > 0$ and let $\widehat{f}_\alpha$ be a global minimizer. Moreover, if $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau(\mathbb{M})} \leq \varrho$ and if the parameter choice*

$$\alpha^* = \varepsilon^{-\frac{u}{u+d/2+\tau}}$$

is chosen such that

$$\mathbf{E}\left[\left\|\widehat{f}_{\alpha^*} - f^\dagger\right\|_{\mathbb{X}}^2\right]^{1/2} = O\left(\varrho^{\frac{2u+d}{2u+2\tau+d}} \varepsilon^{\frac{2\tau}{2\tau+2u+d}}\right) \tag{5.44}$$

Proof. From Lemma 5.7 and Theorem 5.16 with $s = d/2$, we have

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq 2^{\frac{d}{2u}} \widetilde{C}_s^2 \alpha^{-1-\frac{d}{2u}} \varepsilon \|W\|_{B_{2,\infty}^{-d/2}}^2 + c\varrho^2 \alpha^{\frac{\tau}{u}} \tag{5.45}$$

By the parameter choice $\alpha \sim \varrho^{-\frac{2\tau}{2u+2\tau+d}} \varepsilon^{\frac{2\tau}{2\tau+2u+d}}$ we derive

$$\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2 \leq C\varrho^{\frac{4\tau+2d}{2u+2\tau+d}} \varepsilon^{\frac{4\tau}{2\tau+2u+d}} \left(1 + 2^{\frac{s}{u}} \widetilde{C}_s^2 \|W\|_{B_{2,\infty}^{-d/2}}^2\right)$$

Using the large deviation inequality given by Theorem 4.7 with $p = 2$ it follows that

$$\mathbf{E}\left[\|W\|_{B_{2,\infty}^{-d/2}}^2\right] < \infty,$$

then by taking expectation on (5.45) we have

$$\mathbf{E}\left[\left\|\widehat{f}_\alpha - f^\dagger\right\|_{\mathbb{X}}^2\right] \leq C\varrho^{\frac{4\tau+2d}{2u+2\tau+d}} \varepsilon^{\frac{4\tau}{2\tau+2u+d}} \left(1 + 2^{\frac{s}{u}} \widetilde{C}_s^2 \mathbf{E}\left[\|W\|_{B_{2,\infty}^{-d/2}}^2\right]\right) < \infty.$$

Then (5.44) easily follows from the above inequality. $\qquad \square$

Weidling et al [93] also derived order optimal rates for additive random noise under variational source conditions and the help of large deviation inequality, where they analyzed the convergence rates in terms of Besov norms.

As we discussed previously, the regularization parameter α has to be chosen in a proper way, which usually requires knowledge of the index function Ψ. Therefore, in order to adapt to the unknown smoothness of the true solution, a *posteriori* parameter choice rules will be considered. In this kind of parameter choice rule, we do not have to require any knowledge of Ψ. Since the classical discrepancy principle in inverse problems is not implementable due to the continuous setup for stochastic processes, we refer to the Lepskii-type balancing principle for the adaptation.

Theorem 5.19 (A posteriori rates of convergence for Poisson and empirical process data). *Suppose that the assumptions of Theorem 5.17 are satisfied. Morover, if $\widehat{f}_\alpha$ is the global minimizer of (5.2) as well as $\delta > 0$, $a \geq \frac{1}{2}c_1$ or ($a \geq \frac{1}{2}c_2$). Consider the regularization parameters α_j by*

$$\alpha_j = \frac{a\ln(n)}{\sqrt{n}} e^{2j-2}, \quad j \in \mathbb{N}$$

with some number $e > 1$. Then with $m = \min\{j \in \mathbb{N} :\ \alpha_j \geq 1\}$ and the Lepskii-type balancing principle is $\alpha_{bal} = \alpha_{j_{bal}}$ with

$$j_{\text{bal}} := \min \left\{ j \in \{1, \cdots, m\} \middle|\ \left\| \widehat{f}_{\alpha_i} - \widehat{f}_{\alpha_j} \right\|_{\mathbb{X}}^2 \leq 8e^{1-j}\ \text{ for all } i < j \right\}$$

and for $\tilde{d} > d$ with $d \geq 3$, we obtain

$$\mathbf{E}\left[\left\| \widehat{f}_{\alpha_{bal}} - f^\dagger \right\|_{\mathbb{X}}^2 \right]^{1/2} = O\left(\varrho^{\frac{2\tau+\tilde{d}}{2\tau+2u+\tilde{d}}} \left(\frac{\ln(n)}{n} \right)^{\frac{\tau}{2\tau+2u+\tilde{d}}} \right) \quad \text{as } n \to \infty. \qquad (5.46)$$

Moreover, if Conjecture 4.9 is satisfied, then we obtain

$$\mathbf{E}\left[\left\| \widehat{f}_{\alpha_{bal}} - f^\dagger \right\|_{\mathbb{X}}^2 \right] = O\left(\varrho^{\frac{2\tau+d}{2\tau+2u+d}} \left(\frac{\ln(n)}{n} \right)^{\frac{\tau}{2\tau+2u+d}} (\ln n)^p \right) \quad \text{as } n \to \infty. \qquad (5.47)$$

Proof. For brevity, we only consider the convergence results for empirical process data. The proof for the Poisson data is similar to the case of empirical process data. Let n be sufficiently large that the assumptions of Theorem 4.11 (or Theorem 4.12) fulfill with $\eta(n) = a\ln(n)$ if $n \geq \exp(4)$. Consider the events on E_η by

$$E_\eta := \left\{ \| G_n - g^\dagger \|_{B_{2,\infty}^{-\tilde{d}/2}(\mathbb{M})} \leq \frac{C_\rho + \eta(n)}{\sqrt{n}} \right\},$$

which has probability $\mathbf{P}[E_\eta^c] \leq \exp\left(c_1^{-1} \eta(n) \right)$ with constants c_1 and C_{ρ, S_l} given by in Theorem 4.11. As we know from the error decomposition (5.28) and assume $\beta \in [1/2, \infty)$, the total error on the event E_η

$$\left\| \widehat{f}_j - f^\dagger \right\|_{\mathbb{X}} \leq \frac{1}{2} \left(f_1(j) + f_2(j) \right)$$

holds true with

$$f_1(j) = 8\widetilde{C}_{con} 2^{\frac{\tilde{d}}{2u}} \alpha_j^{-\frac{2u-\tilde{d}}{2u}} \cdot \frac{C_\rho + \eta(n)}{\sqrt{n}} = 8\widetilde{C}_{con} 2^{\frac{\tilde{d}}{2u}} \alpha_j^{-\frac{2u-\tilde{d}}{2u}} e^{1-j}$$

and

$$f_2(j) = \Psi(2\alpha_j)^{\frac{1}{2}} = (-\Phi)^* \left(-\frac{1}{2\alpha_j} \right)^{\frac{1}{2}},$$

where the approximation error $f_2(j)$ is typically unknown and monotonically increasing, the data noise error $f_1(j)$ is known and monotonically decreasing if the upper bound is available. If $f^\dagger \in B_{2,\infty}^\tau$, then using Lemma 5.7 we have

$$f_2(j) = C_\Psi \alpha_j^{\frac{\tau}{2u}}$$

for some constant $C_\Psi > 0$. Thus, these functions fulfil the requirement of oracle inequality of Corollary 1 [67] and it is given by

$$\max_{E_\eta} \left\| \widehat{f}_{\alpha_{j_{bal}}} - f^\dagger \right\|_{\mathbb{X}} \leq 3e \min\{ f_1(j) + f_2(j) \mid j \in \{1, \cdots, m\} \}.$$

By plugging the definition of these functions f_1, f_2, we have

$$\max_{E_\eta} \left\| \widehat{f}_{\alpha_{j_{\text{bal}}}} - f^\dagger \right\|_{\mathbb{X}}^2 \le 12e^2 \min \left\{ \widetilde{C}_{con} 2^{\frac{\tilde{d}}{4u}} \frac{(C_\rho + \eta(n))^2}{n\alpha_j} \alpha_j^{-\frac{\tilde{d}}{2u}} + \frac{c\varrho^2}{2} \alpha_j^{\frac{\tau}{u}} \right\}. \tag{5.48}$$

Since $\alpha \mapsto \Psi = (-\Phi)^* \left(-\frac{1}{\alpha} \right)$ is monotonically increasing, we can show that the minimum over $\alpha_1, \cdots, \alpha_m$ can be replaced up to some constant depending only on e by the infimum over $\alpha \ge \alpha_1$ if n is large enough. The sum in (5.48) attains its minimum over $\alpha \in (0, \infty)$ at $\alpha = \alpha_{\text{opt}}$ if and only if the choice $\alpha_{\text{opt}}^{-1} = n^{\frac{2\tau}{2\tau+2u+\tilde{d}}}$ is used. By concavity of Φ we find that

$$\frac{1}{\alpha_{\text{opt}}} \le \lim_{h \searrow 0} \frac{\Phi(\alpha_1) - \Phi(\alpha_1 - h)}{h} \le \frac{\Phi(\alpha_1) - \Phi(t)}{\alpha_1 - t}$$

for all $t \in [0, \alpha_1]$. If we choose $t = 0$, then $\alpha_1/\alpha_{\text{opt}} \le \Phi(\alpha_1)$ for all $n > 0$ and we have $\alpha_1 \le \alpha_{\text{opt}}$ for sufficiently large n. Thus, we can replace the minimum in (5.48) by the infimum over all $\alpha > 0$.

Define the diameter of the subset $\mathcal{B}$ by $\text{diam}(\mathcal{B}) := \sup_{f, \widetilde{f} \in \mathcal{B}} \|f - \widetilde{f}\|_{\mathbb{X}} < \infty$. Then, the reconstruction error in expectation can be bounded as

$$\mathbf{E} \left[\left\| \widehat{f}_{k_{\text{bal}}} - f^\dagger \right\|_{\mathbb{X}}^2 \right] \le \mathbf{P}[E_\eta] \max_{E_\eta} \left\| \widehat{f}_{k_{\text{bal}}} - f^\dagger \right\|_{\mathbb{X}}^2 + \mathbf{P}[E_\eta^c] \max_{E_\eta^c} \left\| \widehat{f}_{k_{\text{bal}}} - f^\dagger \right\|_{\mathbb{X}}^2$$

$$\le C_E \varrho^{\frac{4\tau+2d}{2\tau+2u+\tilde{d}}} \left(\frac{\ln(n)}{n} \right)^{\frac{\tau}{\tau+u+\tilde{d}}} + \exp(c_1^{-1}\eta(n))\text{diam}(\mathcal{B})^2 \tag{5.49}$$

with some constant $C_E = \left(\widetilde{C}_{con}^2 2^{1-\frac{s}{u}} (C_\rho^{(P)} + \eta_h)^2 + c \right) > 0$. Now we bound the second term on the right hand side of above inequality. Since the definition $\eta(n) = a\ln(n)$ and $2ac_1^{-1} \ge 1, \ln(n) \ge 1$, we derive

$$\exp\left(-c_1^{-1}\eta(n) \right) = \left(\frac{1}{n} \right)^{2ac_1^{-1}} \le \left(\frac{\ln(n)}{n} \right)^{2ac_1^{-1}} \le \frac{\ln(n)}{n}$$

$$\le \left(\frac{\ln(n)}{n} \right)^{\frac{4\tau}{2\tau+2u+\tilde{d}}}.$$

By inserting this inequality into (5.49), we obtain (5.46). Finally, (5.47) with $\tilde{d} = d$ easily follows from Conjecture 4.9. □

Chapter 6

The iteratively regularized Gauss-Newton method with additive random noise

In this chapter we will recall the regularization theory for the iteratively regularized Gauss-Newton method and application to the case of additive random noise under variational source conditions. We consider in whole chapter an ill-posed operator equation

$$F(f) = g \tag{6.1}$$

with a nonlinear, Fréchet-differentiable operator $F : \mathcal{B} \subset \mathbb{X} \to \mathbb{Y} = L^2(\mathbb{M})$, where $\mathcal{B}$ is closed domain. Recall the Gaussian white noise model $g^{\mathrm{obs}} = F(f^\dagger) + \varepsilon W$ with stochastic noise level $\varepsilon > 0$.

To solve the operator equation (6.1), we usually replacing the exact operator function $F(f)$ by its linearization. The first order Taylor approximation around a current iterate $\widehat{f}_k$ gives that

$$F(f) \approx F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right).$$

Then the iteratively regularized Gauss-Newton method consists in computing

$$
\widehat{f}_{k+1} \in \operatorname*{argmin}_{f \in \mathcal{B}} \left[\frac{1}{2} \left\| F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right) \right\|_{\mathbb{Y}}^2 \right.
$$
$$
\left. - \left\langle g^{\mathrm{obs}}, F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right) \right\rangle + \frac{\alpha_k}{2} \|f - f_0\|_{\mathbb{X}}^2 \right], \tag{6.2a}
$$

where we also omit the term $\frac{1}{2}\|g^{\mathrm{obs}}\|_{\mathbb{Y}}^2$ since it has no influence on the iteration. The regularization parameters α_k are assumed to satisfy

$$\alpha_0 \leq 1, \quad \lim_{k \to \infty} \alpha_k = 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq C_{dec} \quad \text{for all } k \in \mathbb{N}. \tag{6.2b}$$

This chapter will be contained with some regularization properties of the method (6.2) in the case of additive random noise in the quadratic Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$.

Note that the method (6.2) is not a Newton method in the sense of optimization, in other words, a Newton method applied to $F'[f^*]\left(F(f) - g^{\mathrm{obs}}\right) = 0$, which would require second derivatives of the operator F. Therefore we call (6.2) Gauss-Newton method, which was first introduced by Bakushinskii [3] with $r = 2$. Bauer et al [7] developed this method with the norm powers $r = q = 2$ to the case of stochastic noise based on the spectral theory. As we already discussed in Chapter 5, the convergence results on the Tikhonov regularization for nonlinear operators F hold true

provided Assumption 3.18 is fulfilled. However, the Tikhonov functional

$$f \mapsto \frac{1}{\alpha} \| g^{\text{obs}} - F(f) \|_{\mathbb{Y}}^2 + \frac{1}{2} \| f \|_{\mathbb{X}}^2$$

is not convex for nonlinear operators F and it is difficult to find the global minimizer since the minimization problem is non-convex and (5.2) may have many local minima. To overcome such difficulties, we consider instead the iteratively regularization method (6.2) where we have to solve a convex minimization problem in every iterations.

In section 6.1 we give regularization properties of Gauss-Newton method and present additional assumption to control the nonlinearity error. In section 6.2 we study convergence analysis under a variational source condition as we given in Chapter 3 and we prove convergence and optimal convergence rate results for the iteratively regularized Gauss-Newton method.

6.1 Regularization properties and assumptions

In this section we will shortly review regularization properties of Gauss-Newton method for deterministic noise models with noise level $\| g^{\text{obs}} - F(f^\dagger) \|_{\mathbb{Y}} \leq \delta$ and give some assumptions. The more general form of the iteratively regularized Gauss-Newton method is given by

$$\widehat{f}_{k+1} \in \operatorname*{argmin}_{f \in \mathcal{B}} \left[\left\| F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right) - g^{\text{obs}} \right\|_{\mathbb{Y}}^r + \alpha_k \| f - f_0 \|_{\mathbb{X}}^q \right] \tag{6.3}$$

with some $q, r > 1$, where $f_0 \in \mathcal{B}$ is some initial guess. The regularization parameters a_k are given by (6.2b).

As we already mentioned in Chapter 5 for the basic regularization properties of Tikhonov-type regularization, we also need such kind of regularization properties for the Gauss-Newton method with slightly different prospectives. More precisely, we are interested in the following natural properties:

(a) **Well-posedness**: for any iteration $k \in \mathbb{N}$ and noisy measurement $g^{\text{obs}} \in \mathbb{Y}$ there exists at least single minimizer $\widehat{f}_k$.

(b) **Convergence for exact data**: if there exists some number $K \in \mathbb{N}$ for $g^{\text{obs}} = g^\dagger$, then the iteration $\widehat{f}_K = f^\dagger$. Alternatively speaking, the regularization estimator $\widehat{f}_k$ converges to the solution $f^\dagger$, as $k \to \infty$.

(c) **Convergence for measurement data**: the Newton type estimator $\widehat{f}_K$ converges to the solution $f^\dagger$ as the noise level $\delta \to 0$ and the stopping index K is could chosen in proper way.

Bakushinskii [3] studied the properties of the item (b) and (c) under a Hölder-type source condition of index $\nu \geq \frac{1}{2}$ and proved local convergence result for exact data under a nonlinearity and source condition. Kaltenbacher et al [57] introduced the nonlinearity condition and showed items (b) and (c) without source conditions and for Hölder-type source conditions with index $\nu \in (0, \frac{1}{2})$. Additionally, Hohage [44] studied the convergence rates of the iteratively regularized Gauss-Newton method in the case of logarithmic source conditions (3.7) in Hilbert spaces. Langer and Hohage [61] derive optimal convergence rates under general source conditions. However, these convergence results are obtained based on the spectral theory in

Hilbert spaces and under deterministic error assumption. Furthermore, the convergence rate results for general cases under the source conditions given by multiplicative form are considered by Kaltenbacher and Hofmann [56]. For the stochastic random noise, Bauer et al [7] obtained optimal rates of convergence based on the spectral theory.

To review more information and references on the iteratively regularized Gauss-Newton method we refer to the work of Kaltenbacher et al [56, 57, 58]. More precisely, we simply recall some existing results on convergence rates for the method (6.2) under deterministic setting since it will provide us fundemental tool to achieve the optimal convergence rates for inverse problems with additive random noise. Additionally, the iterative method (6.2) has close connections to other type of iterative methods such as **Levenberg-Marquard method** and **Landweber iteration**. Since it is not our aim to dicuss here, we refer for more details to [56, 57].

Kaltenbacher et al [57] considered the case of an *a priori* stopping rule $K^* = K^*(\delta)$ chosen such that

$$\eta \alpha_{K^*}^{\nu + \frac{1}{2}} \leq \delta < \eta \alpha_k^{\nu + \frac{1}{2}}$$

for all $k \in [0, K^*)$ if the smoothness $\nu \in (0, 1]$. Moreover, if $\nu = 0$, then $\eta \geq \delta \alpha_{K^*}^{-\frac{1}{2}}$ and $K^*(\delta) \to \infty$ as $\delta \to 0$. Furthermore, they show convergence rates by using the discripancy principle, i.e.,

$$\left\| g^{\text{obs}} - F(\widehat{f}_{K^*}) \right\|_{\mathbb{Y}} \leq \tau \delta \leq \| g^{\text{obs}} - F(\widehat{f}_k) \|_{\mathbb{Y}} \tag{6.4}$$

for all $k \in [0, K^*)$, where $\tau > 1$ is some tunning parameter. General norm powers in Banach spaces has been considered by Kaltenbacher et al [58], where they derived convergence rates for (6.2) and (6.3) under *a priori* and *a posteriori* parameter choice rules. But we are only interested in the convergence rate results in the quadratic Hilbert spaces. For more general norm powers as (6.2), we refer to Kaltenbacher et al [58].

In the following, we will present some important nonlinearity assumptions for the analysis of iterative methods (6.2), which will be needed to derive convergence rates. If $\mathcal{S}$ and $\mathcal{T}$ are given by norm powers, we usually use the following tangential cone condition:

Assumption 6.1 (**Classical tangential cone condition**)**.** *For same Banach space* $\tilde{\mathbb{Y}} \subset \mathbb{Y}$, *there exists a constant* $\tilde{\eta} > 0$ *such that*

$$\left\| F(v) - F(u) - F'[u](v - u) \right\|_{\tilde{\mathbb{Y}}} \leq \tilde{\eta} \| F(v) - F(u) \|_{\tilde{\mathbb{Y}}} \tag{6.5}$$

for all $u, v \in \mathcal{B}$.

Note that if Assumption 6.1 holds true, then by the triangle inequality

$$\left\| F'[f^\dagger](f - f^\dagger) \right\|_{\tilde{\mathbb{Y}}} \leq (1 + \tilde{\eta}) \| F(f) - F(f^\dagger) \|_{\tilde{\mathbb{Y}}} \tag{6.6}$$

and we can replace $\| F'[f^\dagger](f - f^\dagger) \|$ by $F(f) - F(f^\dagger)$ only loosing some constant. This is a commonly used assumption introduced by Hanke et al [38] for analyzing regularization methods for nonlinear inverse problems. Alternatively, there exists another frequently used assumption that the Lipschitz continuity of F'. Assume that there exists some constant $L > 0$ such that

$$\left\| F'[u] - F'[v] \right\|_{\mathbb{Y}} \leq L \| u - v \|_{\mathbb{X}}$$

for all $u, v \in \mathcal{B}$. By integration the above inequality implies

$$\|F(v) - F(u) - F'[u](v-u)\|_{\mathbb{Y}} \leq \frac{L}{2}\|v-u\|_{\mathbb{X}}^2 \tag{6.7}$$

for all $u, v \in \mathcal{B}$, which hold true on a sufficiently large set. In practical applications, many operators are smooth such that the condition (6.7) seems to be more useful to control the nonlinearity errors, however, it cannot provide sufficient information for weak source conditions to show convergence rates. In contrast, (6.5) is more restrictive than (6.7) if the operator F is ill-posed, (6.5) could be proven for parameter identification problems in partial differential equations with distributed measurements [24]. However, the conditions (6.5) is very difficult to verify for many important applications in practice. In particular, exponentially ill-posed problems.

6.2 Optimal convergence rates for Gaussian white noise

In this section we present our main results on the convergence rate for Gauss-Newton method with additive Gaussian white noise. Recall from (6.2) that

$$\mathcal{S}(g^{\mathrm{obs}}, g) = \frac{1}{2}\|g\|_{\mathbb{Y}}^2 - \langle g, g^{\mathrm{obs}}\rangle, \tag{6.8}$$

Therefore, by using the error term (3.30) with linearization of g, we have

$$\mathbf{err}(g) := \left\langle \varepsilon W, g - g^\dagger \right\rangle. \tag{6.9}$$

To show the existence of minimizers of (6.2), we need the following lemma.

Lemma 6.2. *Assume that the forward operator $F : \mathcal{B} \subset \mathbb{X} \to B_{2,2}^u(\mathbb{M})$ for $u > d/2$ is Fréchet differentiable and Assumption 5.11 is satisfied. Then the Fréchet derivative $F'[f] : \mathbb{X} \to B_{2,2}^u(\mathbb{M})$ is Hilbert-Schmidt.*

Proof. $F : \mathbb{B} \subset \mathbb{X} \to B_{2,2}^u(\mathbb{M})$ is continuous operator. For any $h > 0$, the first variations

$$F'[f]h := \lim_{t \to 0} \frac{1}{t}\left(F(f + th) - F(f)\right)$$

exists for all $f \in \mathcal{B}$ and $h \in \mathcal{B}$. By the smoothness assumption 5.11 we have

$$\left\|F'[f]\right\|_{B_{2,2}^u(\mathbb{M})} \leq L\|h\|_{\mathbb{X}}$$

which shows boundedness of $F'[f]$ at f. Furthermore, by Proposition 4.8 [93] we find that minimizers of (6.2) exist and $F'[f]$ is compact linear operator. Thus, $F'[f]$ is Hilbert-Schmidt. $\square$

Hohage and Werner [51] studied convergence rates of iteratively regularized Newton-type methods for nonlinear inverse problems with Poisson data, where the reconstruction error bound is given by general Bregman distance. For the additive Gaussian white noise, we will study the convergence rates under the variational source conditions 3.18 with the form (3.38) rather than the multiplicative or additive form of variational inequalities due to keep consistency throughout the thesis.

For brevity, we will use the following notations:

$$d_k := \left\|\widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^2 \quad \text{and} \quad t_k := \left\|F(\widehat{f}_k) - g^\dagger\right\|_{\mathbb{Y}}^2. \tag{6.10}$$

Lemma 6.3. *Let $F : \mathcal{B} \subset \mathbb{X} \to L^\infty(\mathbb{M})$ be a Fréchet differentiable operator such that $\sup_{f \in \mathcal{B}} \|F(f)\|_{L^\infty(\mathbb{M})} < \infty$. Suppose that Assumption 3.18 holds true with $\beta = \frac{1}{2}$ and Assumption 6.1 are satisfied with $\tilde{\mathbb{Y}} = L^2(\mathbb{M})$ and sufficiently small $\tilde{\eta}$. Let the iterates $\widehat{f}_k$ be well-defined as in (6.3) for all $k \in \mathbb{N}$. Then the following inequality*

$$\alpha_{k+1} d_{k+1} + \frac{4}{c_{\tilde{\eta}}} t_{k+1} \le 6\tilde{\eta}^2 t_k + 2\mathbf{err}_{k+1} + 4r_k, \tag{6.11}$$

holds true for $\tilde{\eta} > 0$, where $r_k := \alpha_{k+1}(-\Phi)^* \left(-\frac{1}{c_{\tilde{\eta}}\alpha_{k+1}} \right)$, $c_{\tilde{\eta}} = 8(1 - 4\tilde{\eta}^2)^{-1}$ and

$$\mathbf{err}_{k+1} := \left\langle \varepsilon W, F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - f^\dagger \right) \right\rangle. \tag{6.12}$$

Proof. By the definition of $\widehat{f}_{k+1}$ in (6.3), we have

$$\frac{1}{2} \left\| F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2 -$$
$$\left\langle g^{\mathrm{obs}}, F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\rangle + \frac{\alpha_{k+1}}{2} \left\| \widehat{f}_{k+1} \right\|_{\mathbb{X}}^2$$
$$\le \frac{1}{2} \left\| F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(f^\dagger - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2 -$$
$$\left\langle g^{\mathrm{obs}}, F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(f^\dagger - \widehat{f}_k \right) \right\rangle + \frac{\alpha_{k+1}}{2} \left\| f^\dagger \right\|_{\mathbb{X}}^2$$

for all $k \in \mathbb{N}$. Subtracting $\mathcal{S}(g^{\mathrm{obs}}, g^\dagger) = \frac{1}{2} \|g^\dagger\|_{\mathbb{Y}}^2 - \langle g^{\mathrm{obs}}, g^\dagger \rangle$ on both sides and using Definition 3.11 with $C_{err} = 1$ and g replaced by $F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(f - \widehat{f}_k \right)$ it follows that

$$\left\| g^\dagger - F(\widehat{f}_k) - F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2 + \alpha_{k+1} \left\| \widehat{f}_{k+1} \right\|_{\mathbb{X}}^2$$
$$\le \left\| g^\dagger - F(\widehat{f}_k) - F' \left[\widehat{f}_k \right] \left(f^\dagger - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2 + \alpha_{k+1} \left\| f^\dagger \right\|_{\mathbb{X}}^2 + \mathbf{err}_{k+1} \tag{6.13}$$

for all $k \in \mathbb{N}$, since

$$\left\langle \varepsilon W, F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) - g^\dagger \right\rangle$$
$$- \left\langle \varepsilon W, F(\widehat{f}_k) + F' \left[\widehat{f}_k \right] \left(f^\dagger - \widehat{f}_k \right) - g^\dagger \right\rangle = \left\langle \varepsilon W, F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - f^\dagger \right) \right\rangle$$
$$= \mathbf{err}_{k+1}.$$

From the tangential cone condition 6.1 and using $|a - b|^q \ge 2^{1-q} a^q - b^q$ for $a, b \ge 0$, the first term on the left hand side of (6.13) is bounded by

$$\left\| g^\dagger - F(\widehat{f}_k) - F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2$$
$$\ge \left\| \left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{\mathbb{Y}} - \left\| F(\widehat{f}_{k+1}) - F(\widehat{f}_k) - F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\| \right\|^2$$
$$\ge \frac{1}{2} \left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{\mathbb{Y}}^2 - \left\| F(\widehat{f}_{k+1}) - F(\widehat{f}_k) - F' \left[\widehat{f}_k \right] \left(\widehat{f}_{k+1} - \widehat{f}_k \right) \right\|_{\mathbb{Y}}^2$$
$$\ge \frac{1}{2} \left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{\mathbb{Y}}^2 - \tilde{\eta}^2 \left\| F(\widehat{f}_{k+1}) - F(\widehat{f}_k) \right\|_{\mathbb{Y}}^2$$
$$= \frac{1}{2} \left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{\mathbb{Y}}^2 - \tilde{\eta}^2 \left\| (F(\widehat{f}_{k+1}) - g^\dagger) - (F(\widehat{f}_k) - g^\dagger) \right\|_{\mathbb{Y}}^2$$

$$\geq \left(\frac{1}{2} - 2\tilde{\eta}^2\right) \left\|F(\widehat{f}_{k+1}) - g^\dagger\right\|_{\mathbb{Y}}^2 - 2\tilde{\eta}^2 \left\|F(\widehat{f}_k) - g^\dagger\right\|_{\mathbb{Y}}^2 \tag{6.14}$$

and the first term on the right hand side of (6.13) is also bounded by

$$\left\|g^\dagger - F(\widehat{f}_k) - F'\left[\widehat{f}_k\right]\left(f^\dagger - \widehat{f}_k\right)\right\|_{\mathbb{Y}}^2 \leq \tilde{\eta}^2 \|F(\widehat{f}_k) - g^\dagger\|_{\mathbb{Y}}^2 \tag{6.15}$$

Plugging (6.14) and (6.15) into (6.13) and using (6.10) we obtain

$$\left(\frac{1}{2} - 2\tilde{\eta}^2\right) t_{k+1} + \alpha_{k+1} \left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 \leq 3\tilde{\eta}^2 t_k + \alpha_{k+1} \left\|f^\dagger\right\|_{\mathbb{X}}^2 + \mathbf{err}_{k+1}$$

Setting $\beta = 1/2$ and using the variational source condition 3.18 in the form (3.38) and the above inequalities it follows that

$$\frac{1}{4}d_{k+1} + \frac{1}{4}\left(\frac{1}{2} - 2\tilde{\eta}^2\right)\frac{t_{k+1}}{\alpha_{k+1}} \leq \frac{1}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 - \frac{1}{2}\left\|f^\dagger\right\|_{\mathbb{X}}^2 + \Phi(t_{k+1}) + \frac{1}{4}\left(\frac{1}{2} - 2\tilde{\eta}^2\right)\frac{t_{k+1}}{\alpha_{k+1}}$$

$$\leq \frac{3\tilde{\eta}^2}{2\alpha_{k+1}}t_k + \frac{\mathbf{err}_{k+1}}{2\alpha_{k+1}} + \Phi(t_{k+1}) - \frac{1}{4}\left(\frac{1}{2} - 2\tilde{\eta}^2\right)\frac{t_{k+1}}{\alpha_{k+1}}$$

From the Fenchel conjugate of Φ

$$st_{k+1} + \Phi(t_{k+1}) \leq \sup_{\tau \geq 0}\left[s\tau - (-\Phi)(\tau)\right] = (-\Phi)^*(s)$$

for $s = -(c_{\tilde{\eta}}\alpha_{k+1})^{-1}$ with $c_{\tilde{\eta}} = 8(1 - 4\tilde{\eta}^2)^{-1}$, we derive

$$\frac{1}{4}d_{k+1} + \frac{1}{c_{\tilde{\eta}}\alpha_{k+1}}t_{k+1} \leq \frac{3\tilde{\eta}^2}{2\alpha_{k+1}}t_k + \frac{\mathbf{err}_{k+1}}{2\alpha_{k+1}} + (-\Phi)^*(-\frac{1}{c_{\tilde{\eta}}\alpha_{k+1}}). \tag{6.16}$$

Then rearranging the inequality, we have

$$\alpha_{k+1}d_{k+1} + \frac{4}{c_{\tilde{\eta}}}t_{k+1} \leq 6\tilde{\eta}^2 t_k + 2\mathbf{err}_{k+1} + 4\alpha_{k+1}(-\Phi)^*(-\frac{1}{c_{\tilde{\eta}}\alpha_{k+1}}), \tag{6.17}$$

which yields the assertion. $\qquad\square$

Theorem 6.4 (Error decomposition). *Suppose that Assumptions of Lemma 6.3 hold true. Furthermore, Let Assumptions 5.11 and 6.1 are satisfied with $\tilde{\mathbb{Y}} = B_{2,1}^u(\mathbb{M})$ be applied. Let $\tilde{\eta}, \|\widehat{f}_0 - f^\dagger\|_{\mathbb{X}}, \|F(\widehat{f}_0) - F(f^\dagger)\|_{\mathbb{Y}}$ and $C_{dec} \geq 1$ be sufficiently small. Then there exists some constants $C_1, C_2 > 0$ such that*

$$\left\|\widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^2 \leq C_1\alpha_k^{-1-\frac{d}{2u}}\varepsilon^2 \|W\|_{B_{2,\infty}^{-d/2}}^2 + C_2\Psi(c_{\tilde{\eta}}\alpha_k) \tag{6.18}$$

and

$$\left\|F(\widehat{f}_k) - g^\dagger\right\|_{\mathbb{X}}^2 \leq 2C_1\alpha_k^{-\frac{d}{2u}}\varepsilon^2 \|W\|_{B_{2,\infty}^{-d/2}}^2 + 2C_2\alpha_k\Psi(c_{\tilde{\eta}}\alpha_k) \tag{6.19}$$

for all $k \in \mathbb{N}$.

Proof. We bound $\mathbf{err}_{k+1}$ in Lemma 6.3 by

$$\mathbf{err}_{k+1} = \left\langle \varepsilon W, F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - f^\dagger\right)\right\rangle$$

$$\leq \varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}(\mathbb{M})} \left\| F'[\widehat{f}_k]\left(\widehat{f}_{k+1} - f^\dagger\right)\right\|_{B_{2,1}^{\frac{d}{2}}(\mathbb{M})}. \tag{6.20}$$

By using tangential cone condition 6.1 we have

$$\left\| F'[\widehat{f}_k]\left(\widehat{f}_{k+1} - f^\dagger\right)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))}$$

$$\leq \left\| F'[\widehat{f}_k]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))} + \left\| F'[\widehat{f}_k]\left(\widehat{f}_k - f^\dagger\right)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))}$$

$$\leq \left\| F(\widehat{f}_{k+1}) - F(\widehat{f}_k) - F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))} + \left\| F(\widehat{f}_{k+1}) - F(\widehat{f}_k)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))}$$

$$+ \left\| F(f^\dagger) - F(\widehat{f}_k) - F'\left[\widehat{f}_k\right]\left(f^\dagger - \widehat{f}_k\right)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))} + \left\| F(\widehat{f}_k) - F(f^\dagger)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))}$$

$$\leq (1+\tilde{\eta})\left\| F(\widehat{f}_{k+1}) - F(f^\dagger)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))} + 2(1+\tilde{\eta})\left\| F(\widehat{f}_k) - F(f^\dagger)\right\|_{B_{2,1}^{\frac{d}{2}}((\mathbb{M}))}. \tag{6.21}$$

To bound (6.21), we need to apply the interpolation theorem 4.5. For $u > \frac{d}{2}$ there exist $C_I > 0$ such that

$$\left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{B_{2,1}^{\frac{d}{2}}(\mathbb{M})} \leq C_I \left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{B_{2,2}^{u}(\mathbb{M})}^{\frac{d}{2u}} \left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{B_{2,2}^{0}(\mathbb{M})}^{1-\frac{d}{2u}}$$

$$= C_I \left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{B_{2,2}^{u}(\mathbb{M})}^{\frac{d}{2u}} \left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{L^2(\mathbb{M})}^{1-\frac{d}{2u}}. \tag{6.22}$$

Furthermore, we still need to bound the first factor in (6.22). By the smoothness assumption 5.11, we have

$$\left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{B_{2,2}^{u}(\mathbb{M})}^{\frac{d}{2u}} \leq L \left\| \widehat{f}_{k+1} - f^\dagger\right\|_{\mathbb{X}}^{\frac{d}{2u}} \tag{6.23}$$

for some Lipschitz constant $L > 0$. Similarly,

$$\left\| F(\widehat{f}_k) - g^\dagger\right\|_{B_{2,1}^{\frac{d}{2}}(\mathbb{M})} \leq C_I L \left\| \widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^{\frac{d}{2u}} \left\| F(\widehat{f}_k) - g^\dagger\right\|_{L^2(\mathbb{M})}^{1-\frac{d}{2u}}, \tag{6.24}$$

Finally, inserting (6.21), (6.22) and (6.23) into (6.20) and using (6.10) we derive

$$\mathbf{err}_{k+1} \leq (1+\tilde{\eta})C_I L \varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}} \left\| \widehat{f}_{k+1} - f^\dagger\right\|_{\mathbb{X}}^{\frac{d}{2u}} \left\| F(\widehat{f}_{k+1}) - g^\dagger\right\|_{\mathbb{Y}}^{1-\frac{d}{2u}}$$

$$+ 2(1+\tilde{\eta})C_I L \varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}} \left\| \widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^{\frac{d}{2u}} \left\| F(\widehat{f}_k) - g^\dagger\right\|_{\mathbb{Y}}^{1-\frac{d}{2u}}$$

$$= (1+\tilde{\eta})C_I L \varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}(\mathbb{M})} \left(d_{k+1}^{\frac{d}{4u}} t_{k+1}^{\frac{1}{2}-\frac{d}{4u}} + 2d_k^{\frac{d}{4u}} t_k^{\frac{1}{2}-\frac{d}{4u}} \right)$$

$$= (1+\tilde{\eta})C_I L \varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}(\mathbb{M})} \alpha_{k+1}^{-\frac{d}{4u}} \left((\alpha_{k+1} d_{k+1})^{\frac{d}{4u}} t_{k+1}^{\frac{1}{2}-\frac{d}{4u}} + 2(\alpha_{k+1} d_k)^{\frac{d}{4u}} t_k^{\frac{1}{2}-\frac{d}{4u}} \right)$$

$$= R_{k+1}(\alpha_{k+1} d_{k+1})^{\frac{d}{4u}} t_{k+1}^{\frac{1}{2}-\frac{d}{4u}} + 2R_{k+1}(\alpha_{k+1} d_k)^{\frac{d}{4u}} t_k^{\frac{1}{2}-\frac{d}{4u}}, \tag{6.25}$$

where $R_{k+1} := (1+\tilde{\eta})C_I L\varepsilon \|W\|_{B_{2,\infty}^{-\frac{d}{2}}(\mathbb{M})} \alpha_{k+1}^{-\frac{d}{4u}}$. By applying Young's inequality

$$(\theta a)\left(\frac{b}{\theta}\right) \leq \frac{1}{p}(\theta a)^p + \frac{1}{q}\left(\frac{b}{\theta}\right)^q$$

with $p = \frac{4u}{d}, \theta = \left(\frac{2u}{d}\right)^{d/4u}$ and $a = \alpha_{k+1}d_{k+1}, b = R_{k+1}t_{k+1}^{\frac{1}{2}-\frac{d}{4u}}$, then we have

$$R_{k+1}(\alpha_{k+1}d_{k+1})^{\frac{d}{4u}}t_{k+1}^{\frac{1}{2}-\frac{d}{4u}} \leq \frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{4u-d}{4u}\left(\left(\frac{2u}{d}\right)^{-\frac{d}{4u-d}}R_{k+1}^{\frac{4u}{4u-d}}\right)t_{k+1}^{\frac{2u-d}{4u-d}}. \quad (6.26)$$

Using Young's inequality once more on the right hand side of (6.26), we derive

$$\frac{4u-d}{4u}\left(\left(\frac{2u}{d}\right)^{-\frac{d}{4u-d}}R_{k+1}^{\frac{4u}{4u-d}}\right)t_{k+1}^{\frac{2u-d}{4u-d}}$$

$$\leq \frac{2}{c_{\tilde{\eta}}}t_{k+1} + \left(\frac{2}{c_{\tilde{\eta}}}\frac{4u-d}{4u-2d}\right)^{-\frac{2u-d}{2u}}\left(\frac{2u}{d}\right)^{-\frac{d}{4u-d}}R_{k+1}^2,$$

where we choose $p = \frac{4u-d}{2u-d}, \theta = \left(\frac{2}{c_{\tilde{\eta}}}\frac{4u-d}{4u-2d}\right)^{\frac{2u-d}{4u-d}}$ and $a = t_{k+1}^{\frac{2u-d}{4u-d}}$.

Plugging this inequality into (6.26) and setting $C_d := \left(\frac{2}{c_{\tilde{\eta}}}\frac{4u-d}{4u-2d}\right)^{-\frac{2u-d}{2u}}\left(\frac{2u}{d}\right)^{-\frac{d}{4u-d}}$, we have

$$R_{k+1}(\alpha_{k+1}d_{k+1})^{\frac{d}{4u}}t_{k+1}^{\frac{1}{2}-\frac{d}{4u}} \leq \frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{2}{c_{\tilde{\eta}}}t_{k+1} + C_d R_{k+1}^2. \quad (6.27)$$

Analogously, by using Young's inequality for the second term on the right hand side of (6.25), we derive

$$2R(\alpha_{k+1}d_k)^{\frac{d}{4u}}t_k^{\frac{1}{2}-\frac{d}{4u}} \leq \frac{1}{4}\alpha_{k+1}d_k + \left(\frac{1}{c_{\tilde{\eta}}} - 6\tilde{\eta}^2\right)t_k + C_e R_{k+1}^2. \quad (6.28)$$

for some constant C_e. Plugging (6.27) and (6.28) into (6.25), we derive

$$\mathbf{err}_{k+1} \leq \frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{1}{4}\alpha_{k+1}d_k + \frac{2}{c_{\tilde{\eta}}}t_{k+1} + \left(\frac{1}{c_{\tilde{\eta}}} - 6\tilde{\eta}^2\right)t_k + CR_{k+1}^2 \quad (6.29)$$

Then inserting inequality (6.29) into (6.11) and rearranging we obtain

$$\frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{2}{c_{\tilde{\eta}}}t_{k+1} \leq \frac{1}{4}\alpha_{k+1}d_k + \frac{1}{c_{\tilde{\eta}}}t_k + 4r_k + CR_{k+1}^2. \quad (6.30)$$

Define $\Delta_{k+1} := \frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{2}{c_{\tilde{\eta}}}t_{k+1}$. Then we have the following recursive inequality

$$\Delta_{k+1} \leq \frac{1}{2}\Delta_k + 4r_k + CR_{k+1}^2. \quad (6.31)$$

The concavity of $\Phi^{1+\epsilon}$ for some $\epsilon > 0$ implies that $\Phi(Ct) \leq C^{\frac{1}{1+\epsilon}}\Phi(t)$ for all $C \geq 1$ and $t \geq 1$. Using the substitution $\frac{t}{C} = C^{1/\epsilon}\eta$ yields

$$(-\Phi)^*\left(-\frac{1}{C\alpha}\right) = \sup_{\eta \geq 0}\left[-C^{\frac{1}{\epsilon}}\frac{\eta}{\alpha} + \Phi\left(C^{\frac{1+\epsilon}{\epsilon}}\eta\right)\right] \leq C^{\frac{1}{\epsilon}}(-\Phi)^*\left(-\frac{1}{\alpha}\right).$$

From (6.2b) and the above inequality we obtain

$$r_k \leq C_{\mathrm{dec}} \alpha_{k+2} (-\Phi)^* \left(-\frac{1}{C_{\mathrm{dec}} c_{\tilde{\eta}} \alpha_{k+2}} \right) \leq C_{\mathrm{dec}}^{\frac{1+\epsilon}{\epsilon}} r_{k+1}, \tag{6.32}$$

Now we introduce $\gamma_{nl} := C_{\mathrm{dec}}^{\frac{1+\epsilon}{\epsilon}} \left(1 - \frac{1}{2} C_{\mathrm{dec}}^{\frac{1+\epsilon}{\epsilon}} \right)^{-1}$. Then we apply an induction method to prove

$$\Delta_k \leq \gamma_{nl}(4r_k + CR_k^2), \tag{6.33}$$

for $k = 0, 1, 2, \cdots$. For $k = 0$, we specify the second smallness assumption by $t_0 \leq \gamma_{nl} r_0$. Note that γ_{nl} is well-defined and positive if $C_{\mathrm{dex}} > 1$ is sufficiently small. Suppose that (6.33) holds true for some k. Then by (6.31) and (6.32) we have

$$\begin{aligned}
\Delta_{k+1} &\leq \frac{1}{2}\Delta_k + 4r_k + CR_{k+1}^2 \\
&\leq \left(\frac{1}{2}\gamma_{nl} + 1 \right)(4r_k + CR_{k+1}^2) \\
&\leq C_{\mathrm{dec}}^{\frac{1+\epsilon}{\epsilon}} \left(\frac{1}{2}\gamma_{nl} + 1 \right)(4r_{k+1} + CR_{k+2}^2) \\
&= \gamma_{nl}(4r_{k+1} + CR_{k+2}^2),
\end{aligned}$$

which completes the proof of (6.33). Thus, plugging (6.33) into the right hand side of (6.30) and replace α_{k+1} by α_k yields the assertion. More precisely, by non-negativity of d_k and t_k and omitting the second term in the above inequality, we derive (6.18). Similarly, by omitting the first term in the above inequality, we obtain (6.19). $\qquad \square$

The deterministic error bound in Theorem 6.4 is the essential step for our convergence analysis in the case of stochastic noise data and we are able to achieve optimal rates of convergence for additive Gaussian white noise. As we know, Werner [94] obtained optimal convergence rates for iteratively regularized Newton-type method of ill-posed inverse problems under Hölder-type and logarithmic source conditions in Theorem 6.11 and Theorem 6.14, but their results restricted for deterministic error assumptions. We are now able to prove optimal convergence rates for mildly ill-posed inverse problems with additive Gaussian white noise.

Theorem 6.5 (Optimal convergence rates). *Suppose that Assumptions of Theorem 6.4 are satisfied. If η and the initial value of $\|g^\dagger - F(\widehat{f}_0)\|_\mathbb{Y}$ are sufficiently small and if $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau(\mathbb{M})} \leq \varrho$ and the iterates $\widehat{f}_k$ are defined by (6.3) with index $k \in \mathbb{N}$ and the iteration is stopped at*

$$k^* = \min \left\{ k \in \mathbb{N} \,\middle|\, \alpha_k = \left(\frac{\varepsilon}{\varrho} \right)^{\frac{\tau}{u+d/2+\tau}} \right\}, \tag{6.34}$$

then we obtain the following convergence rate

$$\mathbf{E}\left[\left\| \widehat{f}_{k^*} - f^\dagger \right\|_\mathbb{X}^2 \right]^{1/2} = O\left(\varrho^{\frac{2\tau+d}{2\tau+2u+d}} \varepsilon^{\frac{2\tau}{2\tau+2u+d}} \right)$$

as $\varepsilon \to 0$.

Proof. From Lemma 5.7 we find that $f^\dagger$ satisfies the varitional source conditions with some index function

$$\Psi(\alpha_k) = cc_{\tilde{\eta}} \varrho^2 \alpha_k^{\frac{\tau}{u}}$$

for a constant $c > 0$ if $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau} \leq \varrho$. Then by combining the results of Theorem 6.4 with (6.18), we get

$$\left\|\widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^2 \leq 2C_{con}\alpha_k^{-1-\frac{d}{2u}}\varepsilon^2\|W\|_{B_{2,\infty}^{-d/2}(\mathbb{M})}^2 + cc_{\tilde{\eta}}\varrho^2\alpha_k^{\frac{\tau}{u}}.$$

With the parameter choice rule $\alpha_{k^*} \sim \left(\frac{\varepsilon}{\varrho}\right)^{\frac{2\tau}{u+d/2+\tau}}$, we have

$$\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2 \leq \left(2C_{con}\|W\|_{B_{2,\infty}^{-d/2}(\mathbb{M})} + cc_{\tilde{\eta}}\right)\varrho^{\frac{4\tau+2d}{2\tau+2u+d}}\varepsilon^{\frac{4\tau}{\tau+u+d/2}}. \tag{6.35}$$

Finally, by the deviation inequality in Theorem 4.7 we derive

$$\mathbf{P}\left(\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2 \geq (C+r)\varrho^{\frac{4\tau+2d}{2\tau+2u+d}}\varepsilon^{\frac{4\tau}{\tau+u+d/2}}\right) \leq \exp\left(-\frac{r^2}{4}\right) \tag{6.36}$$

where C, r are positive constants independent of $f^\dagger, \widehat{f}_k$ and ε. Taking expectation on (6.35) with (4.10) we have

$$\mathbf{E}\left[\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2\right] \leq \left(2C_{con}\mathbf{E}\left[\|W\|_{B_{2,\infty}^{-d/2}(\mathbb{M})}\right] + cc_{\tilde{\eta}}\right)\varrho^{\frac{4\tau+2d}{2\tau+2u+d}}\varepsilon^{\frac{4\tau}{\tau+u+d/2}} < \infty,$$

which yields the assertion. $\qquad\square$

Chapter 7

Iteratively regularized Newton-type method with Poisson and empirical process data

In this final chapter we again consider ill-posed inverse problems with nonlinear operator equations

$$F(f) = g,$$

where $F : \mathcal{B} \subset \mathbb{X} \to \mathbb{Y}$ be a nonlinear forward operator between the Hilbert spaces $\mathbb{X}$ and $\mathbb{Y}$ and Fréchet-differentiable. To solve such nonlinear operator equations, we already discussed the iteratively regularized Gauss-Newton method in previous chapter. In this chapter, we particularly study Newton-type methods for the Kullback-Leibler type data fidelity and quadratic penalty functionals in Hilbert spaces. This leads to

$$\widehat{f}_{k+1} \in \underset{f \in \mathcal{B}}{\operatorname{argmin}} \left[\mathcal{S}\left(F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right), G_n \right) + \frac{\alpha}{2} \left\| f - \widehat{f}_0 \right\|_{\mathbb{X}}^2 \right], \qquad \text{(7.1a)}$$

where $\widehat{f}_0 \in \mathcal{B}$ is some initial value, G_n denotes the observed data, $\mathcal{S}$ is data fidelity term and the regularization parameters $\alpha_k > 0$ satisfying

$$\alpha_0 \leq 1, \quad \lim_{k \to \infty} \alpha_k = 0, \quad 1 \leq \frac{\alpha_k}{\alpha_{k+1}} \leq C_{dec} \quad \text{for all } k \in \mathbb{N}. \qquad \text{(7.1b)}$$

The method (7.1) is called Newton-type method, which is obviously seen as a generalization of the iteratively-regularized Gauss-Newton method. Our aim in this chapter is to derive optimal convergence rates for iteratively regularized Newton-type method under variational source condition 3.18. To the best of our knowledge, no optimal order of convergence rates exist for the iteratively regularized Newton-type method of mildly ill-posed inverse problems with Poisson and empirical process data under some noise assumptions. Although Werner and Hohage [51, 53], Dunker and Hohage [24] studied the statistical convergence analysis for the iteratively regularized Newton-type method (7.1) with Poisson and empirical process data respectively in Banach space setting, they derived suboptimal rates of convergence for the Newton-type methods. Therefore, we will attempt to improve their convergence bound and ideally derive optimal convergence rates for the method (7.1) in the Hilbert space setting. In Section 7.1 we will recall nonlinearity conditions on the operator F. The known convergence rate results for Poisson and empirical process data are given in Section 7.2. Finally, improved convergence rates of the expectation of the reconstruction error for Poisson and empirical process data will be derived with an explicit order as the noise level tends to zero in Section 7.3.

7.1 Nonlinearity conditions

To derive an optimal bound on the reconstruction error, the first step is to bound the noise error $\mathbf{err}_k$ in probability, where $\mathbf{err}_k$ was defined as the sum of two error terms $\mathbf{err}(g)$ with g replaced by the first order Taylor expansion of operator equation $F(f)$ around $\widehat{f}_k$ evaluated at $f^\dagger$ and $\widehat{f}_k$ respectively. Recall the error term

$$\mathbf{err}(g) := \begin{cases} \int_{\mathbb{M}} \ln\left(\frac{g+\zeta}{g^\dagger+\zeta}\right)(\mathrm{d}G_n - g^\dagger \mathrm{d}x), & g \geq -\frac{\zeta}{2}, \\ \infty, & \text{else.} \end{cases} \tag{7.2}$$

The main purpose here is to show the noise error (7.2) are finite and sufficiently small for all $g \in \mathbb{Y}$. To show the convergence of the iteration (7.1), we need some conditions on it and to ensures that the first order Taylor approximation of data fidelity functional $\mathcal{S}(g^{\mathrm{obs}}; F(\widehat{f}_k) + F'[\widehat{f}_k](f - \widehat{f}_k))$ is a sufficiently approximate good to $\mathcal{S}(g^{\mathrm{obs}}; F(f))$. Since in the linearized form of $F(f)$ if cannot be guaranteed that the functions $F(\widehat{f}_k) + F'[\widehat{f}_k](f - \widehat{f}_k)$ are non-negative, we consider the data fidelity functional $\mathcal{S}$ with offset parameter $\zeta > 0$.

As we discussed in Chapter 6, the nonlinearity conditions (6.5) and (6.7) are also not available for the non-quadratic data fidelity functional $\mathcal{S}$, where $\mathcal{S}$ does not necessarily fulfill a triangle inequality. Therefore, we need a generalized tangential cone condition to control the nonlinearity error, which was introduced first time by Hanke et al [38]. In the case of general source conditions (3.4), the item (b) in chapter 6 implies a tangential cone condition and no convergence rates results exist under weaker nonlinearity conditions with general source conditions φ.

Assumption 7.1 (Generalized tangential cone condition). *Let $\mathcal{T}$ be the data fidelity functionals for exact data. There exists a constant η (assume it is sufficiently small) and $C_{tc} \geq 1$ such that*

$$\frac{1}{C_{tc}}\mathcal{T}(g^\dagger; F(\tilde{f})) - \eta\mathcal{T}(g^\dagger; F(f)) \leq \mathcal{T}(g^\dagger; F(f) + F'[f](\tilde{f} - f))$$
$$\leq C_{tc}\mathcal{T}(g^\dagger; F(\tilde{f})) + \eta\mathcal{T}(g^\dagger; F(f)) \tag{7.3}$$

for all $f, \tilde{f} \in \mathcal{B}$.

Remark 7.2. *In addition, Hohage and Werner [51] also derive the error bound under the generalized tangential cone condition with the noise data fidelity functionals $\mathcal{S}$. Similarly, Werner [94] additionally studied convergence rates for iteratively regularized Newton-type methods under a Lipschitz type of condition and pointed out that both conditions can ensure the nonlinearity of F fits together with the data fidelity functional $\mathcal{S}$.*

Werner and Hohage in Lemma 5.2 [51] discussed the relation between Assumption 7.1 and Assumption 6.1 that Assumption 7.1 follows from the classical tangential cone condition 6.1 if $\mathcal{T}$ is given by the norm powers.

Lemma 7.3. *Assume that $\mathcal{S}(g_1, g_2) = \|g_1 - g_2\|_{\mathbb{Y}}^q$. If the operator F fulfills the classical tangential cone condition 6.1 with $\widetilde{\eta} \geq 0$ sufficiently small, then general tangential cone condition 7.1 are satisfied with $\eta = 2^{2q-2}\widetilde{\eta}^q$ and constant C_{tc} is given by*

$$C_{tc} = \max\left\{\frac{1}{2^{1-q} - 2^{q-1}\widetilde{\eta}^q}, 2^{q-1} + \widetilde{\eta}^q 2^{2q-2}\right\} \geq 1. \tag{7.4}$$

Proof. See Lemma 5.2 [51]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

This condition is valid for linear operator F as $\eta = 0$ and $C_{tc} = 1$.

7.2 Known convergence rates under variational source conditions

In this section we will shortly review convergence rates results for iteratively regularized Newton-type methods of inverse problems with Poisson and empirical process data derived by Hohage and Werner [51, 53] and Dunker and Hohage [24]. They also studied convergence analysis for nonlinear ill-posed inverse problems with Poisson and empirical process data. First of all, we present convergence results of (7.1) for a *priori* stopping rule.

Theorem 7.4 (A priori stopping rule). *Suppose that Assumptions 2.10 and 7.1 with (b) and Assumption 3.18 are satisfied and the operator $F : \mathcal{B} \to H^2(\mathbb{M})$ such that*

$$\sup_{f, \widetilde{f} \in \mathcal{B}} \left\| F(f) + F^{'}[f](f - \widetilde{f}) \right\|_{H^s(\mathbb{M})} < \infty$$

is fulfilled with smoothness index $s > d/2$. Moreover, η and $\mathbf{KL}(g^\dagger, F(\widehat{f}_0))$ are supposed to be sufficiently small. If we choose the stopping index k_* by

$$k_* = \min \left\{ k \in \mathbb{N} \,\middle|\, \frac{1}{\alpha_k} \geq -\inf \partial \left(-\Phi \left(\frac{1}{\sqrt{n}} \right) \right) \right\}, \tag{7.5}$$

we derive

$$\mathbf{E} \left[\left\| \widehat{f}_{k_*} - f^\dagger \right\|_{\mathbb{X}}^2 \right] = O \left(\Phi \left(\frac{1}{\sqrt{n}} \right) \right), \tag{7.6}$$

as $n \to \infty$.

Proof. See Theorem 7.3 [94]. $\qquad\square$

As we discussed in Chapter 5, this convergence rate is shown to be suboptimal. Under the assumptions of Lemma 5.7 and with parameter choice

$$\alpha_{k_*} \sim (t\varrho^4)^{-\frac{u}{2\tau + 2u}},$$

the convergence rate (7.6) can be written

$$\mathbf{E} \left[\left\| \widehat{f}_{k_*} - f^\dagger \right\|_{\mathbb{X}}^2 \right] = O \left(\varrho^{\frac{4u}{2\tau + 2u}} \left(\frac{1}{n} \right)^{\frac{\tau}{2\tau + 2u}} \right), \tag{7.7}$$

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 (7.7) is not satisfied the order optimality of convergence rates for mildly ill-posed inverse problems.

Since the aforementioned stopping rule is not available for requiring knowledge of the index function Φ in (3.35) characterizing smoothness of the unknown solution $f^\dagger$, they studied a *a posteriori* parameter choice rules, in that case this stopping rule does not require knowledge of Φ. Again, the well-known discrepancy principle cannot be applied due to the stochastic setting of the Poisson and empirical processes. If the true solution $f^\dagger$ fulfill the source conditions 3.18 for the index functions Ψ, then

by the Lepskii type balancing principle the convergence in expectation

$$\mathbf{E}\left[\left\|\widehat{f}_{k_{\mathrm{bal}}} - f^{\dagger}\right\|_{\mathbb{X}}^{2}\right] \to 0$$

as $n \to \infty$, see [53]. Unfortunately, these convergence results in expectation for the iteratively regularized Newton-type method also shown to be not optimal. However, it is unknown under which conditions a better convergence result holds true for variational source conditions in Hilbert spaces so far. This is our main purpose to improve such reconstruction error bounds and ideally show optimal convergence rates of iteratively regularized Newton-type method of inverse problems with Poisson and empirical process data in the next section.

7.3 Improved convergence rates for Poisson and empirical process data

In this section we will study the convergence analysis for the iteratively regularized Newton-type method of mildly ill-posed problems with Poisson and empirical process data under stochastic noise assumptions. In order to improve the convergence bound for the iteratively regularized Newton-type method with stochastic data, it is an essential step to bound up the propagated error term (7.2) and we first need to bound the logarithm of the denisty function in terms of the Besov norm $B_{2,1}^{s}(\mathbb{M})$. For simplicity, we use the following notations:

$$d_k := \|\widehat{f}_k - f^{\dagger}\|_{\mathbb{X}}^{2}, \quad t_k := \mathcal{T}_{\zeta}(F(\widehat{f}_k), g^{\dagger}) \tag{7.8}$$

Lemma 7.5. *Let $F : \mathcal{B} \subset \mathbb{X} \to L^{\infty}(\mathbb{M})$ be a Fréchet differentiable operator such that $\sup_{f \in \mathcal{B}} \|F(f)\|_{B_{2,1}^{s}(\mathbb{M})} < \infty$ with some $s \in [d/2, u)$. Suppose that Assumption 3.18 holds true with $\beta = \frac{1}{2}$ and Assumption 7.1 is satisfied and η is sufficiently small. Let the iterates $\widehat{f}_k$ be well-defined as in (7.1) for all $k \in \mathbb{N}$. Then the inequality*

$$\alpha_{k+1} d_{k+1} + \frac{1}{C_{tc}} t_{k+1} \leq 4\eta t_k + 2\mathbf{err}_{k+1} + \lambda_k, \tag{7.9}$$

holds true for $\eta > 0$, where $\lambda_k := 2\alpha_{k+1}(-\Phi)^*(-\frac{1}{2C_{tc}\alpha_{k+1}})$ and

$$
\begin{aligned}
\mathbf{err}_{k+1} := \; & \mathbf{err}\left(F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right) \\
& - \mathbf{err}\left(F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f^{\dagger} - \widehat{f}_k\right)\right)
\end{aligned}
\tag{7.10}
$$

Proof. By definition of $\widehat{f}_{k+1}$ in (7.1) $F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right)$ and setting $f = f^{\dagger}$, we have

$$
\begin{aligned}
& \mathcal{S}_{\zeta}\left(G_n, F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right) + \frac{\alpha_{k+1}}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^{2} \\
& \leq \mathcal{S}_{\zeta}\left(G_n, F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f^{\dagger} - \widehat{f}_k\right)\right) + \frac{\alpha_{k+1}}{2}\left\|f^{\dagger}\right\|_{\mathbb{X}}^{2}
\end{aligned}
$$

for all $k \in \mathbb{N}$. Subtracting $\mathcal{S}(G_n, g^{\dagger}) = \int_{\mathbb{M}} g^{\dagger} \mathrm{d}x - \int_{\mathbb{M}} \ln g^{\dagger} \mathrm{d}G_n$ on both sides and using Definition 3.11 with $C_{err} = 1$ and g replaced by $F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f - \widehat{f}_k\right)$ it

follows that

$$
\begin{aligned}
\mathcal{T}_\zeta &\left(g^\dagger, F(\widehat{f}_k) - F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right) + \frac{\alpha_{k+1}}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 \\
&\leq \mathcal{T}_\zeta\left(g^\dagger, F(\widehat{f}_k) - F'\left[\widehat{f}_k\right]\left(f^\dagger - \widehat{f}_k\right)\right) + \frac{\alpha_{k+1}}{2}\left\|f^\dagger\right\|_{\mathbb{X}}^2 + \mathbf{err}_{k+1}
\end{aligned} \tag{7.11}
$$

for all $k \in \mathbb{N}$. From general tangential cone condition 7.1, (7.11) is bounded by

$$
\begin{aligned}
\frac{1}{C_{tc}} &\mathcal{T}_\zeta(g^\dagger, F(\widehat{f}_{k+1})) - \eta\mathcal{T}_\zeta(g^\dagger, F(\widehat{f}_k)) + \frac{\alpha_{k+1}}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 \\
&\leq C_{tc}\mathcal{T}_\zeta(g^\dagger, F(f^\dagger)) + \eta\mathcal{T}_\zeta(g^\dagger, F(\widehat{f}_k)) + \frac{\alpha_{k+1}}{2}\left\|f^\dagger\right\|_{\mathbb{X}}^2 + \mathbf{err}_{k+1},
\end{aligned} \tag{7.12}
$$

where $\mathcal{T}_\zeta(g^\dagger, F(f^\dagger)) = 0$. Using the notations (7.8), then (7.12) is written as

$$
\frac{1}{C_{tc}}t_{k+1} + \frac{\alpha_{k+1}}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 \leq 2\eta t_k + \frac{\alpha_{k+1}}{2}\left\|f^\dagger\right\|_{\mathbb{X}}^2 + \mathbf{err}_{k+1} \tag{7.13}
$$

for all $k \in \mathbb{N}$. Setting $\beta = 1/2$ and using the variational source condition 3.18 and (7.13) it follows that

$$
\begin{aligned}
\frac{1}{2}d_{k+1} + \frac{t_{k+1}}{2C_{tc}\alpha_{k+1}} &\leq \frac{1}{2}\left\|\widehat{f}_{k+1}\right\|_{\mathbb{X}}^2 - \frac{1}{2}\left\|f^\dagger\right\|_{\mathbb{X}}^2 + \Phi(t_{k+1}) + \frac{t_{k+1}}{2C_{tc}\alpha_{k+1}} \\
&\leq \frac{2\eta}{\alpha_{k+1}}t_k + \frac{\mathbf{err}_{k+1}}{\alpha_{k+1}} + \Phi(t_{k+1}) - \frac{t_{k+1}}{2C_{tc}\alpha_{k+1}}
\end{aligned}
$$

for all $k \in \mathbb{N}$. From the Fenchel conjugation of Φ

$$
st_{k+1} + \Phi(t_{k+1}) \leq \sup_{\tau \geq 0}\left[s\tau - (-\Phi)(\tau)\right] = (-\Phi)^*(s)
$$

for $s = -(2C_{tc}\alpha_{k+1})^{-1}$, we derive

$$
\frac{1}{2}d_{k+1} + \frac{1}{2C_{tc}\alpha_{k+1}}t_{k+1} \leq \frac{2\eta}{\alpha_{k+1}}t_k + \frac{\mathbf{err}_{k+1}}{\alpha_{k+1}} - (-\Phi)^*\left(-\frac{1}{2C_{tc}\alpha_{k+1}}\right).
$$

Then rearranging the inequality yields the assertion. $\qquad\square$

Theorem 7.6 (Error bounds). *Suppose that the assumptions of Lemma 7.5 hold true for $s \geq d/2$. Moreover, let Assumptions 2.10 and 6.1 hold true with $\widetilde{\mathbb{Y}} = B_{2,1}^u(\mathbb{M})$. Let the iterates $\widehat{f}_k$ be well-defined as in (7.1). If η and initial residual $\|\widehat{f}_0 - f^\dagger\|_{\mathbb{X}}$ and $\mathcal{T}_\zeta(g^\dagger, F(\widehat{f}_0))$ are sufficiently small, then there exists some constant $\widetilde{C}_\zeta > 0$ such that*

$$
\left\|\widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^2 \leq \widetilde{C}_\zeta\alpha_k^{-1-\frac{s}{u}}\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^{-s}(\mathbb{M})}^2 + \Psi(C_{tc}\alpha_k) \tag{7.14}
$$

and

$$
\mathcal{T}_\zeta(F(\widehat{f}_k), g^\dagger) \leq 2\widetilde{C}_\zeta\alpha_k^{-1-\frac{s}{u}}\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^{-s}(\mathbb{M})}^2 + 2\Psi(C_{tc}\alpha_k) \tag{7.15}
$$

holds true for all $k \in \mathbb{N}$.

Proof. Bu Hölder's inequality and Theorem 4.6 and with the help of Lemma 5.10, and the interpolation inequality the error (7.10) in Lemma 7.5 can be bounded by

$$
\mathbf{err}_{k+1} = \mathbf{err}\left(F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right)\right) - \mathbf{err}\left(F(\widehat{f}_k) + F'\left[\widehat{f}_k\right]\left(f^\dagger - \widehat{f}_k\right)\right)
$$

$$= \int_{\mathbb{M}} \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{g^\dagger + \zeta} \right) \left(dG_n - g^\dagger dx \right)$$

$$- \int_{\mathbb{M}} \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta}{g^\dagger + \zeta} \right) \left(dG_n - g^\dagger dx \right)$$

$$= \int_{\mathbb{M}} \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \left(dG_n - g^\dagger dx \right)$$

$$\leq \|G_n - g^\dagger\|_{B_{2,\infty}^{-s}(\mathbb{M})} \left\| \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{B_{2,1}^{s}(\mathbb{M})}$$

$$\leq C C_1 \|G_n - g^\dagger\|_{B_{2,\infty}^{-s}(\mathbb{M})} \left\| \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{B_{2,2}^{u}(\mathbb{M})}^{\frac{s}{u}}$$

$$\times \left\| \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{B_{2,2}^{0}(\mathbb{M})}^{1 - \frac{s}{u}} \tag{7.16}$$

for some constant $C, C_1 > 0$ and $u > s > 0$, which can be derived as we discussed in Lemma 5.10. As a consequence of Lemma 5.13 and by Theorem 2.8.3 in Triebel [89] for the bounded Lipschitz domain $\mathbb{M}$, the first factor on the right hand side of (7.16) is bounded by

$$\left\| \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{B_{2,2}^{u}(\mathbb{M})}$$

$$\leq C_2 \left(\frac{2\|g^\dagger\|_{L^\infty} + \zeta}{\zeta} \right)^m$$

$$\times \left\| \frac{F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger)}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right\|_{B_{2,2}^{u}(\mathbb{M})}. \tag{7.17}$$

and the second factor is also bounded by

$$\left\| \ln \left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{L^2(\mathbb{M})}$$

$$\leq C_2 \left(\frac{2\|g^\dagger\|_{L^\infty} + \zeta}{\zeta} \right)$$

$$\times \left\| \frac{F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger)}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right\|_{L^2(\mathbb{M})}. \tag{7.18}$$

From the variational characterization of the iterate $\widehat{f}_{k+1}$ it follows that

$$F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(\widehat{f}_{k+1} - \widehat{f}_k\right) \geq -\frac{\zeta}{2}.$$

and from the tangential cone condition 7.1 we find that

$$F\left(\widehat{f}_k\right) + F'\left[\widehat{f}_k\right]\left(f^\dagger - \widehat{f}_k\right) \geq -\frac{\zeta}{2}.$$

the above inequalities imply

$$\left\| \ln\left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{B_{2,2}^u(\mathbb{M})} \leq C_{\zeta,1} \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{B_{2,2}^u(\mathbb{M})},$$

where $C_{\zeta,1} = C_2 \left(\frac{2\|g^\dagger\|_{L^\infty} + \zeta}{\zeta} \right)^m$ and the second factor is also bounded by

$$\left\| \ln\left(\frac{F(\widehat{f}_k) + F'[\widehat{f}_k](\widehat{f}_{k+1} - \widehat{f}_k) + \zeta}{F(\widehat{f}_k) + F'[\widehat{f}_k](f^\dagger - \widehat{f}_k) + \zeta} \right) \right\|_{L^2(\mathbb{M})} \leq C_{\zeta,2} \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{L^2(\mathbb{M})},$$

where $C_{\zeta,2} = C_2 \left(\frac{2\|g^\dagger\|_{L^\infty} + \zeta}{\zeta} \right)$. Finally, the noise error (7.16) is can be rewritten as follows

$$\begin{aligned}
\mathbf{err}_{k+1} \leq {} & C_\zeta \|G_n - g^\dagger\|_{B_{2,\infty}^{-s}(\mathbb{M})} \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{B_{2,2}^u(\mathbb{M})}^{\frac{s}{u}} \\
& \times \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{L^2(\mathbb{M})}^{1 - \frac{s}{u}}
\end{aligned} \tag{7.19}$$

where $C_\zeta = C_{\zeta,1} C_{\zeta,2} C C_1$. From (6.21) and smoothness assumption 5.11, we have

$$\begin{aligned}
\left\| F'[\widehat{f}_k]\left(\widehat{f}_{k+1} - f^\dagger \right) \right\|_{B_{2,2}^u(\mathbb{M})} \leq {} & (1 + \tilde{\eta}) \left\| F(\widehat{f}_{k+1}) - F(f^\dagger) \right\|_{B_{2,2}^u(\mathbb{M})} \\
& + 2(1 + \tilde{\eta}) \left\| F(\widehat{f}_k) - F(f^\dagger) \right\|_{B_{2,2}^u(\mathbb{M})},
\end{aligned} \tag{7.20}$$

and

$$\begin{aligned}
\left\| F'[\widehat{f}_k]\left(\widehat{f}_{k+1} - f^\dagger \right) \right\|_{L^2(\mathbb{M})} \leq {} & (1 + \tilde{\eta}) \left\| F(\widehat{f}_{k+1}) - F(f^\dagger) \right\|_{L^2(\mathbb{M})} \\
& + 2(1 + \tilde{\eta}) \left\| F(\widehat{f}_k) - F(f^\dagger) \right\|_{L^2(\mathbb{M})},
\end{aligned} \tag{7.21}$$

Then, the smoothness assumption 5.11 implies that

$$\left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{B_{2,2}^u(\mathbb{M})} \leq L \left\| \widehat{f}_{k+1} - f^\dagger \right\|_{\mathbb{X}} \tag{7.22}$$

for some Lipschitz constant $L > 0$. By using Lemma 5.12, the second factor on the right hand side of (7.21) can be bounded by

$$\begin{aligned}
\left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{L^2(\mathbb{M})} \leq {} & \left(\frac{2}{3} \|g^\dagger\|_{L^\infty(\mathbb{M})} + \frac{4}{3} \|F(\widehat{f}_k)\|_{L^\infty(\mathbb{M})} + 2\zeta \right)^{1/2} \\
& \times \mathcal{T}_\zeta(F(\widehat{f}_{k+1}), g^\dagger)^{1/2},
\end{aligned} \tag{7.23}$$

where $\sup_{f \in \mathcal{B}} \|F(\widehat{f}_k)\|_{L^\infty(\mathbb{M})} < \infty$ such that (7.23) is bounded by

$$\left\| F(\widehat{f}_{k+1}) - g^\dagger \right\|_{L^2(\mathbb{M})} \leq C_{g^\dagger} \mathcal{T}_\zeta(F(\widehat{f}_{k+1}), g^\dagger)^{1/2}. \tag{7.24}$$

with

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

Plugging (7.22) into (7.20), we have

$$\left\| F'[\widehat{f}_k] \left(F(\widehat{f}_{k+1}) - f^\dagger \right) \right\|_{B_{2,2}^u(\mathbb{M})} \le L(1+\eta) \left\| \widehat{f}_{k+1} - f^\dagger \right\|_{\mathbb{X}}$$
$$+ 2L(1+\eta) \left\| \widehat{f}_k - f^\dagger \right\|_{\mathbb{X}} \qquad (7.25)$$

Similarly, by inserting (7.24) into (7.21) we have

$$\left\| F'[\widehat{f}_k] \left(F(\widehat{f}_{k+1}) - f^\dagger \right) \right\|_{L^2(\mathbb{M})} \le C_1(1+\eta)\mathcal{T}_\zeta(F(\widehat{f}_{k+1}), g^\dagger)^{1/2}$$
$$+ 2C_1(1+\eta)\mathcal{T}_\zeta(F(\widehat{f}_k), g^\dagger)^{1/2}. \qquad (7.26)$$

Finally, by inserting (7.25) and (7.26) into (7.19) and using notations (7.8), we derive

$$\mathbf{err}_{k+1} \le C_\zeta \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}(\mathbb{M})} \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{B_{2,2}^u(\mathbb{M})}^{\frac{s}{u}} \left\| F'[\widehat{f}_k](\widehat{f}_{k+1} - f^\dagger) \right\|_{B_{2,2}^0(\mathbb{M})}^{1-\frac{s}{u}}$$

$$\le C_2(1+\eta) \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}(\mathbb{M})} \left(d_{k+1}^{\frac{s}{2u}} t_{k+1}^{\frac{1}{2}-\frac{s}{2u}} + 2d_k^{\frac{s}{2u}} t_k^{\frac{1}{2}-\frac{s}{2u}} \right)$$

$$= C_2(1+\eta) \left\| G_n - g^\dagger \right\|_{B_{2,\infty}^{-s}(\mathbb{M})} \alpha_{k+1}^{-\frac{s}{2u}} \left((\alpha_{k+1}d_{k+1})^{\frac{s}{2u}} t_{k+1}^{\frac{1}{2}-\frac{s}{2u}} \right.$$

$$\left. + 2(\alpha_{k+1}d_k)^{\frac{s}{2u}} t_k^{\frac{1}{2}-\frac{d}{4u}} \right)$$

$$= R_k(\alpha_{k+1}d_{k+1})^{\frac{s}{2u}} t_{k+1}^{\frac{1}{2}-\frac{s}{2u}} + 2R_k(\alpha_{k+1}d_k)^{\frac{s}{2u}} t_k^{\frac{1}{2}-\frac{s}{2u}}, \qquad (7.27)$$

where $R_k := C_2(1+\eta)\|G_n - g^\dagger\|_{B_{2,\infty}^{-s}(\mathbb{M})} \alpha_{k+1}^{-\frac{s}{2u}}$.

The rest of the proof is quite similar with Gaussian white noise case. Therefore, using Young's inequality for both terms on the right hand side of (7.27), we obtain

$$\mathbf{err}_k \le \frac{1}{4}\alpha_{k+1}d_{k+1} + \frac{1}{8}\alpha_{k+1}d_k + \frac{1}{4C_{tc}}t_{k+1} + \left(\frac{1}{8C_{tc}} - 2\eta \right) t_k + C_3 R_k^2. \qquad (7.28)$$

for some constant $C_3 > 0$. Then by plugging (7.28) into (7.9) and rearranging, we obtain

$$\frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{1}{2C_{tc}}t_{k+1} \le \frac{1}{4C_{tc}}t_k + \frac{1}{4}\alpha_{k+1}d_k + \lambda_k + 2C_3 R_k^2 \qquad (7.29)$$

Define $\Delta_{k+1} := \frac{1}{2}\alpha_{k+1}d_{k+1} + \frac{1}{2C_{tc}}t_{k+1}$. Then we have the following recursive inequality

$$\Delta_{k+1} \le \frac{1}{2}\Delta_k + \lambda_k + 2C_3 R_k^2. \qquad (7.30)$$

From (6.32) we know that $\lambda_k \le C_{\text{dec}}^{\frac{1+\epsilon}{\epsilon}} \lambda_{k+1}$. Here we will use

$$\gamma_{nl} := C_{\text{dec}}^{\frac{1+\epsilon}{\epsilon}} \left(1 - \frac{1}{2}C_{\text{dec}}^{\frac{1+\epsilon}{\epsilon}} \right)^{-1}.$$

Then we can prove the following inequality by an induction.

$$\Delta_k \le \gamma_{nl}(\lambda_k + 2C_3 R_k^2). \qquad (7.31)$$

The rest of the proof can be done as in the proof of Theorem 6.4. $\qquad \square$

Now will show improved rates of convergence for mildly ill-posed inverse problems with Poisson and empirical process data simultaneously, these main results can be summarized as following:

Theorem 7.7 (Improved rates for Poisson and empirical process data). *Suppose that the assumptions of Theorem 7.6 hold true. Moreover, if η and the initial value of $\|\widehat{f}_0 - f^\dagger\|_{\mathbb{X}}$ and $\mathcal{T}_\zeta(g^\dagger, F(\widehat{f}_0))$ are sufficiently small and if $f^\dagger \in B_{2,\infty}^\tau(\mathbb{M})$ for some $\tau \in (0, u)$ with $\|f^\dagger\|_{B_{2,\infty}^\tau(\mathbb{M})} \leq \varrho$ and the iterates $\widehat{f}_k$ is defined by (7.1) and the iteration is stopped at*

$$k^* = \min\left\{ k \in \mathbb{N} \,\middle|\, \frac{1}{\alpha_k} = \left(\frac{1}{\sqrt{n}\varrho}\right)^{\frac{u}{u+\tilde{d}/2+\tau}} \right\}$$

with the index $s = \tilde{d}/2$ for $\tilde{d} > d$, then we derive

$$\mathbf{E}\left[\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2\right]^{1/2} = O\left(\varrho^{\frac{2\tau+\tilde{d}}{2\tau+2u+\tilde{d}}} \left(\frac{1}{n}\right)^{\frac{2\tau}{2\tau+2u+\tilde{d}}}\right), \tag{7.32}$$

as $n \to \infty$. Moreover, if Conjecture 4.9 holds true for $\tilde{d} = d$, then we obtain

$$\mathbf{E}\left[\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2\right]^{1/2} = O\left(\varrho^{\frac{2\tau+d}{2\tau+2u+d}} \left(\frac{1}{n}\right)^{\frac{2\tau}{2\tau+2u+d}} (\ln n)^p\right), \tag{7.33}$$

as $n \to \infty$.

Proof. From Lemma 5.7, we know that

$$\Psi(\alpha) = C\varrho^2 \alpha^{\frac{\tau}{u}}$$

for some constant $C > 0$. Therefore, by monotoncity of Ψ, we have

$$\Psi(C_{tc}\alpha_k) \leq (C_{tc})^{\frac{1}{\epsilon}}\Psi(\alpha_k) = C\varrho^2(C_{tc})^{\frac{1}{\epsilon}}\alpha_k^{\frac{\tau}{u}}. \tag{7.34}$$

Then by combining the error bound (7.14) in Theorem 7.6 together with (7.34), we have

$$\left\|\widehat{f}_k - f^\dagger\right\|_{\mathbb{X}}^2 \leq \widetilde{C}_\zeta^2 \alpha_k^{-1-\frac{s}{u}}\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^s(\mathbb{M})}^2 + C(C_{tc})^{\frac{1}{\epsilon}}\alpha_k^{\frac{\tau}{u}}.$$

By the parameter choice $k^* \sim (\varrho\sqrt{n})^{-\frac{u}{u+\tilde{d}/2+\tau}}$ with $s = \tilde{d}/2$ for $\tilde{d} > d$ implies that

$$\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2 \leq \left(\widetilde{C}_\zeta^2 n\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^s(\mathbb{M})}^2 + C(C_{tc})^{\frac{1}{\epsilon}}\right)\varrho^{\frac{4\tau+2\tilde{d}}{2\tau+2u+\tilde{d}}} \left(\frac{1}{n}\right)^{\frac{2\tau}{2\tau+2u+\tilde{d}}} \tag{7.35}$$

Then, by taking expectation on both side of (7.35) and Theorem 4.11 and 4.12 we obtain

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

if $\tilde{d} > d$ or if Conjecture 4.9 holds true with $p = 0$, then

$$\mathbf{E}\left[\left\|\widehat{f}_{k^*} - f^\dagger\right\|_{\mathbb{X}}^2\right] \leq \left(\widetilde{C}_\zeta^2\mathbf{E}\left[n\left\|G_n - g^\dagger\right\|_{B_{2,\infty}^s(\mathbb{M})}^2\right] + C(C_{tc})^{\frac{1}{\epsilon}}\right)\varrho^{\frac{4\tau+2\tilde{d}}{2\tau+2u+\tilde{d}}} \left(\frac{1}{n}\right)^{\frac{2\tau}{2\tau+2u+\tilde{d}}}$$

$$< \infty$$

which implies (7.32) and (7.33) for $p = 0$. If $p > 0$ and $\tilde{d} = d$, then

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

and we derive (7.33). □

As we have discussed in Chapter 5, the convergence rates of Theorem 7.7 for the iteratively regularized Newton-type method with Poisson and empirical process data are well improved with the help of the large deviation inequalities in negative Besov norms. By comparing the convergence rates in Theorem 7.4 and Theorem 7.7 we find that the convergence rates in Theorem 7.7 are close to the optimal rates. If the Conjecture 4.9 holds true with $p = 0$, then the convergence rates that we obtained in Theorem 7.7 are order optimal. However, it is still open question to obtain optimal order for the Newton type method with Poisson and empirical process data. The optimal convergence rates for iteratively regularized Gauss-Newton method with additive Gaussian white noise have been derived in Chapter 6. Moreover, by comparing the convergence results for (7.1) with the results for Tikhonov-type regularization in Chapter 5, we additionally need a nonlinearity assumption 7.1 to control the nonlinearity of F fit together with data fidelity functional.

We are now closing this thesis with some remarks. In this thesis we mainly focused on the convergence rates for variational regularization and iteratively regularized Newton-type methods of mildly ill-posed inverse problems with stochastic data, where stochastic noise processes described by exponential families. In particular, we studied the statistical convergence analysis for variational and iterative methods with Poisson and empirical process data and considerably improved reconstruction error bounds by using some interpolation approaches and applying the large deviation inequalities for Poisson and empirical processes in negative Besov norms. However, we are currently not able to show our conjecture 4.9 as an essential motivation. Therefore, we will leave this goal for future research, which we believe to be of independent interest.

Bibliography

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill-posed problems. *Inverse Problems*, 10(6):1217–1229, 1994.

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

[3] A. B. Bakushinskii. The problem of the convergence of the iteratively regularized Gauss-Newton method. *Comput. Math. Math. Phys.*, 32(9):1353–1359, 1992.

[4] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *J. Mach. Learn. Res.*, 6:1705–1749, 2005.

[5] H. T. Banks and K. Kunisch. *Estimation Techniques for Distributed Parameter Systems*. Birkhauser Boston, Inc., Secaucus, NJ, USA, 1989.

[6] J. Bardsley. A theoretical framework for the regularization of Poisson likelihood estimation problems. *Inverse Problems and Imaging*, 4(1):11–17, 2010.

[7] F. Bauer, T. Hohage, and A. Munk. Iteratively regularized Gauss-Newton method for nonlinear inverse problems with random noise. *SIAM Journal on Numerical Analysis*, 47(3):1827–1846, 2009.

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

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

[10] N. Bissantz, T. Hohage, and A. Munk. Consistency and rates of convergence of nonlinear tikhonov regularization with random noise. *Inverse Problems - INVERSE PROBL*, 20:1773–1789, 2004.

[11] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numerical Analysis*, 45:2610–2636, 2007.

[12] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.

[13] J. M. Borwein and A. S. Lewis. Convergence of best entropy estimates. *SIAM Journal on Optimization*, 1(2):191–205, 1991.

[14] G. Bourdaud and W. Sickel. Composition operators on function spaces with fractional order of smoothness. *RIMS Kokyuroku Bessatsu*, 26:93–132, 2011.

[15] R. I. Boţ and B. Hofmann. An extension of the variational inequality approach for obtaining convergence rates in regularization of nonlinear ill-posed problems. *Journal of Integral Equations and Applications*, 22(3):369–392, 2010.

[16] M. Burger, J. Flemming, and B. Hofmann. Convergence rates in ℓ^1-regular - ization if the sparsity assumption fails. *Inverse Problems*, 29(2):025013, 2013.

[17] M. Burger, T. Helin, and H. Kekkonen. Large noise in variational regularization. *Transactions of Mathematics and Its Applications*, 2(1):1–45, 2018.

[18] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411–1421, 2004.

[19] 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.

[20] F. Comte, V. Genon-Catalot, and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007.

[21] I. Csiszar. I-divergence geometry of probability distributions and minimization problems. *The Annals of Probability*, 3(1):146–158, 1975.

[22] I. Daubechies, M. Defrise, and C. D. Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[23] D. L. Donoho. Nonlinear solution of linear inverse problems by Wavelet - vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2(2):101–126, 1995.

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

[25] S. Efromovich. Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. *IEEE Transactions on Information Theory*, 43(4):1184–1191, 1997.

[26] P. P. B. Eggermont. Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM Journal on Mathematical Analysis*, 24(6):1557–1576, 1993.

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

[28] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Springer Netherlands, 1996.

[29] H. W. Engl and J. Zou. A new approach to convergence rate analysis of Tikhonov regularization for parameter identification in heat conduction. *Inverse Problems*, 16(6):1907–1923, 2000.

[30] S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Problems*, 18(4):R55–R97, 2002.

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

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

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

[34] D. Fournier, L. Gizon, M. Holzke, and T. Hohage. Pinsker estimators for local helioseismology: inversion of travel times for mass-conserving flows. *Inverse Problems*, 32(10):105002, 2016.

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

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

[37] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Dover, 1952.

[38] M. Hanke, A. Neubauer, and O. Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numerische Mathematik*, 72(1):21–37, 1995.

[39] S. Hell and E. H. K. Stelzer. Properties of a 4Pi confocal fluorescence microscope. *Journal of the Optical Society of America A*, 9(12):2159, 1992.

[40] S. W. Hell. Far-field optical nanoscopy. *Science*, 316(5828):1153–1158, 2007.

[41] M. Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79(1):135–163, 1999.

[42] 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(3):987, 2007.

[43] B. Hofmann and M. Yamamoto. On the interplay of source conditions and variational inequalities for nonlinear ill-posed problems. *Applicable Analysis*, 89(11):1705–1727, 2010.

[44] T. Hohage. Logarithmic convergence rates of the iteratively regularized Gauss - Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems*, 13(5):1279, 1997.

[45] T. Hohage. *Iterative Methods in Inverse Obstacle Scattering: Regularization Theory of Linear and Nonlinear Exponentially Ill-Posed Problems*. PhD thesis, 1999.

[46] T. Hohage. Regularization of exponentially ill-posed problems. *Numerical Functional Analysis and Optimization*, 21(3-4):439–464, 2000.

[47] T. Hohage and P. Miller. Optimal convergence rates for sparsity promoting Wavelet - regularization in Besov spaces. *Inverse Problems*, 2019.

[48] T. Hohage and M. Pricop-Jeckstadt. Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise. *Inverse Problems and Imaging*, 2:271–290, 2008.

[49] T. Hohage and F. Weidling. Verification of a variational source condition for acoustic inverse medium scattering problems. *Inverse Problems*, 31:075006, 2015.

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

[51] 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.

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

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

[54] O. Ivanyshyn and R. Kress. Identification of sound-soft 3D obstacles from phaseless data. *Inverse Problems and Imaging*, 4(1):131–149, 2010.

[55] K. M. Johannes Kerstan and J. Mecke. *Infinitely divisible point processes*. Wiley series in probability and mathematical statistics, 1978.

[56] B. Kaltenbacher and B. Hofmann. Convergence rates for the iteratively regularized Gauss - Newton method in Banach spaces. *Inverse Problems*, 26(3):035007, 2010.

[57] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Walter de Gruyter, 2008.

[58] B. Kaltenbacher, F. Schöpfer, and T. Schuster. Iterative methods for nonlinear ill-posed problems in Banach spaces: convergence and applications to parameter identification problems. *Inverse Problems*, 25(6):065003, 2009.

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

[60] Y. A. Kutoyants. *Statistical Inference for Ergodic Diffusion Processes*. Springer London, 2004.

[61] S. Langer and T. Hohage. Convergence analysis of an inexact iteratively regularized Gauss-Newton method under general source conditions. *Journal of Inverse and Ill-posed Problems*, 15(3):311–327, 2007.

[62] M. Ledoux and M. Talagrand. *Probability in Banach Spaces*. Springer Berlin Heidelberg, 1991.

[63] O. V. Lepskii. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability & Its Applications*, 35(3):454–466, 1991.

[64] B. A. Mair and F. H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM Journal on Applied Mathematics*, 56(5):1424–1444, 1996.

[65] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28(2):863–884, 2000.

[66] P. Math and S. V. Pereverzev. Geometry of linear ill-posed problems in variable Hilbert scales. *Inverse Problems*, 19(3):789–803, 2003.

[67] P. Mathé. The Lepskii principle revisited. *Inverse Problems*, 22(3):L11–L15, 2006.

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

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

[70] D. W. Nychka and D. D. Cox. Convergence rates for regularized solutions of integral equations from discrete noisy data. *The Annals of Statistics*, 17(2):556–572, 1989.

[71] F. O'Sullivan. A statistical perspective on ill-posed inverse problems. *Statistical Science*, 1(4):502–518, 1986.

[72] O. Papaspiliopoulos, Y. Pokern, G. O. Roberts, and A. M. Stuart. Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531, 2012.

[73] M. S. Pinsker. Optimal filtering of square-integrable signals in Gaussian noise. *Probl. Peredachi Inf*, 16(2):52–68, 1980.

[74] Y. Pokern, A. Stuart, and J. van Zanten. Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. *Stochastic Processes and their Applications*, 123(2):603–628, 2013.

[75] C. Pöschl. *Tikhonov Regularization with General Residual Term*. PhD thesis, 2008.

[76] 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.

[77] H. Risken. *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer Berlin Heidelberg, 1989.

[78] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational Methods in Imaging*. Springer New York, 2009.

[79] H. J. Schmeisser and H. Triebel. *Topics in Fourier Analysis and Function Spaces, 1st Edition*. Wiley, 1987.

[80] Z. Schuss. *Theory and Applications of Stochastic Processes*. Springer New York, 2010.

[81] T. Schuster, B. Kaltenbacher, B. Hofmann, and K. S. Kazimierski. *Regularization Methods in Banach Spaces*. DE GRUYTER, 2012.

[82] V. G. Spokoiny. Adaptive drift estimation for nonparametric diffusion model. *The Annals of Statistics*, 28(3):815–836, 2000.

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

[84] R. Stück, M. Burger, and T. Hohage. The iteratively regularized Gauss–Newton method with convex constraints and applications in 4Pi microscopy. *Inverse Problems*, 28(1):015012, 2011.

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

[86] M. Talagrand. New concentration inequalities in product spaces. *Inventiones Mathematicae*, 126(3):505–563, 1996.

[87] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Math. Dokl.*, 4:1035–1038, 1963.

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

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

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

[91] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *Journal of the American Statistical Association*, 80(389):8–20, 1985.

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

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

[94] F. Werner. *Inverse problems with Poisson data: Tikhonov-type regularization and iteratively regularized Newton methods*. PhD thesis, 2012.

[95] 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.

[96] P. Wojtaszczyk. *A Mathematical Introduction to Wavelets*. Cambridge University Press, 1997.

[97] Z.-B. Xu and G. F. Roach. Characteristic inequalities of uniformly convex and uniformly smooth Banach spaces. *Journal of Mathematical Analysis and Applications*, 157(1):189–210, 1991.

[98] C. A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. *Inverse Problems*, 25(2):025006, 2009.

Curriculum Vitae

Name	Yusufu Simayi (aka Yusup Ismayil)
Address	Institute for Numerical and Applied Mathematics University of Göttingen Lotzestraße 16-18, 37083, Göttingen / Germany
Email:	y.simayi@math.uni-goettingen.de

Personal details:

Gender	Male
Date of birth	1st of October, 1984
Place of birth	Aksu city, Xinjiang, China
Citizenship	Chinese (Uyghur)

Education:

02/2016– 08/2019	Ph.D student of mathematics at the University of Göttingen (Germany) Supervisor: Prof. Dr. Thorsten Hohage
10/2012– 01/2016	Research assistant of statistics at the University of Konstanz (Germany) Supervisor: Prof. Dr. Jan Beran
09/2008– 06/2011	Master student of Applied Mathematics at the University of Xinjiang (China) Supervisor: Prof. Dr. Geni Gupur Thesis title: *Other eigenvalues of the M/M/1 retrial queuing model with special retrial times*
09/2002– 06/2008	Bachelor student of Applied Mathematics at the University of Xinjiang (China) including the preparation years (09/2002–07/2-2004)
09/1999–07/2002	3rd high school student at Bay county of Xinjiang

Research Experience:

02/2016– 08/2019	DFG scholarship holder of the research training Group 2088:Discovering structure in complex data: Statistics meets Optimization and Inverse Problems
10/2012–12/2014	DFG scholarship holder of the research grant BE2123/11-1: Copulas and spectral measures: Statistical modelling under long-range dependence
01/2015– 06/2015	Scholarship for the outstanding self-finance Xinjiang students