Convergence Rates and Uncertainty Quantification for Inverse Problems

Hanne Kekkonen

Academic dissertation

To be presented for public examination with the permission of the Faculty of Science of the University of Helsinki in Auditorium CK112 in the Kumpula Campus (Gustaf Hällströmin katu 2b, Helsinki) on 26th August 2016 at 12 o'clock.

Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

HELSINKI 2016

ISBN 978-951-51-2373-2 (paperback)
ISBN 978-951-51-2374-9 (PDF)

http://ethesis.helsinki.fi

Unigrafia Oy
Helsinki 2016

Acknowledgements

First and foremost I want to thank my advisor Matti Lassas for introducing me to the world of inverse problems and patiently explaining things as many times as needed. I am deeply grateful to him for pointing me in the right direction and letting me find the answers.

I am also indebted to my second advisor Tapio Helin for all his time and for the opportunities I have gotten mainly because of him. I want to express my gratitude to Samuli Siltanen for his help with computational questions and for the freedom to tinker and play with the 3D printer whenever my mind needed a break from mathematical problems.

I wish cordially to thank my pre-examiners Professor Stefan Kindermann and Professor Shuai Lu and my opponent Professor Christian Clason for their valuable time spent reading my thesis. I would further like to thank Professor Martin Burger for mathematical collaboration and the whole group in Münster for their hospitality which always made Münster feel like a second home department for me.

The atmosphere at the Department of Mathematics and Statistics of the University of Helsinki made it a very nice and inspiring place to work. I am grateful to the whole extended inverse problems lunch group for all the interesting mathematical and the slightly less scientific conversations during lunch and coffee breaks: Paola, Teemu, Andreas, Martina and all the rest. You made the choice of doing a PhD seem like a sane one. A special thanks to Cliff for delightful tea breaks and improving the English of my thesis in many parts including the acknowledgements (excluding this sentence).

To my parents who always encouraged and supported me in my endeavours and happily fed me long after I should have learned to fill my own fridge: Kiitos!

Lastly, I thank the Emil Aaltonen Foundation, the Academy of Finland and the Finnish Center of Excellence in Inverse Problems Research for financial support which made my work possible.

Warwick, July 2016

Hanne Kekkonen

The thesis consists of this overview and the following articles:

Publications

[I] H. KEKKONEN, M. LASSAS AND S. SILTANEN, *Analysis of regularized inversion of data corrupted by white Gaussian noise*, Inverse Problems, 30(4):045009, 2014.

[II] M. BURGER, T. HELIN AND H. KEKKONEN, *Large Noise in Variational Regularization*, arXiv:1602.00520.

[III] H. KEKKONEN, M. LASSAS AND S. SILTANEN, *Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators*, Inverse Problems, 32(8):085005, 2016.

Author's contribution to the publications

[I] Theoretical analysis and most of the writing are due to the author.

[II] The author made major contributions to the theoretical analysis and writing of the paper.

[III] The major ideas, analysis and the writing are due to the author.

Contents

1 Introduction

We often encounter situations where it is quite difficult to interpret uniquely our observations of the world. These observations can be seen as incomplete measurement data but luckily we usually also have a priori information about the circumstances. If you ask your colleague over coffee ‘How are you?’ and get a rather ambiguous ‘Hmph’ in answer it might be hard to decide if the person is tired, grumpy or just Finnish. On the other hand, if it is early Monday morning you automatically add this regularizing knowledge to your deduction and might be more inclined to conclude that they are just tired. Interpreting the mood of a person from their expressions and gestures can be seen as an example of an every day inverse problem.

Similarly in mathematics inverse problems arise from the need to get information from indirect measurements of an unknown object of interest. As opposed to a direct problem, where the causes are known and one wants to know the effect, in inverse problems the effect is given and we want to recover the cause. For example in X-ray tomography the direct problem is, if the internal structure of a physical body is known, then determine the X-ray projection images. The corresponding inverse problem is to reconstruct the inner structure of a patient from the X-ray projection images. Another example is image processing where the inverse problem is to produce a sharp image from a blurred or noisy photograph.

The forward problems are well-posed and they are usually numerically stable and can be solved reliably. Inverse problems on the other hand are mathematically difficult to solve and are characterized by extreme sensitivity to measurement noise and modelling errors. A problem is called ill-posed or an inverse problem if it breaks at least one of the following conditions for well-posedness as defined by Jacques Hadamard [34],

(i) Existence: there should be at least one solution.

(ii) Uniqueness: there should be at most one solution.

(iii) Stability: the solution must depend continuously on data.

Difficulties with existence and uniqueness can often be overcome by mathematical reformulation of the problem. For numerical inverse problems violation of condition (iii) is usually the one that causes most problems. The lack of stability means that even small noise in the data can cause an arbitrary large error in the solution. This makes finding the true solution impossible in practice. Instead one tries to find a reasonably good estimate for the unknown.

A classical method of uncovering such an approximated solution is to use regularization. In this approach the original problem is modified by introducing additional information, usually in the form of a penalty functional, to make the problem stable. Regularization is essentially a trade-off between fitting the data and reducing the penalty term. This will of course introduce new error to the method and hence we want to keep the modification as small as possible. These methods are efficient in practice and have been studied in depth. The research of regularization methods remains active and there exist many excellent books on the topic, see for instance [25, 44, 65, 76]. In classical regularization theory the unknown and the noise are assumed to be deterministic and the magnitude of the noise is assumed to be known and small. Regularization methods are designed so that the regularized solution converges to the true solution when the noise level goes to zero.

Another approach to solving inverse problems is to view them from a statistical point of view [14]. Statistical modelling of the measurement error allows studying a wider range of noise behaviour and makes more sense in many applications [26]. Statistical inverse problems can be divided in two groups, Bayesian and frequentist. The main difference between the two is the interpretation of the concept of probability. The Bayesian idea is that probability is a quantity that represents a state of subjective knowledge or belief. Frequentists on the other hand see probability as a frequency of a phenomenon over time.

In the Bayesian approach we model the unknown quantity and the noise as random variables. This means we have to assign probability distribution for both of them. In many applications the statistics of the noise can be determined quite well and modelled accurately. All the information we have about the object of interest before making a measurement is coded to a prior distribution. One of the core difficulties of Bayesian inversion is describing the prior knowledge in the form of a probability distribution. When the measurement is done we update our prior information to a posterior distribution using the measurement model and the Bayes formula. Hence the solution to a Bayesian inverse problem is a probability distribution. However, an approximated solution is often given as a point estimate. We can, for example, study the mean or the mode of the distribution. The posterior distribution also offers the possibility of uncertainty quantification and assessing the reliability of the point estimates.

In contrast to the Bayesian paradigm, frequentists assume that the unknown is deterministic and only the noise is random. There are various statistical estimation techniques that can be used in such settings. In frequentist approach one often aims to find an estimator that minimizes a specific risk function [4, 18]. A popular estimator in frequentist statistics is the maximum likelihood estimate which is the solution that has most likely produced the measured data. The problem with maximum likelihood estimation is that it does not take into account the instability of inverse problems. Even though the pure frequentists perceive the unknown to be deterministic sometimes a less strict approach is employed. For example, one can assign a prior that includes the true unknown and then apply the Bayes formula. Next one assumes that the unknown is a fixed realization from the prior, giving a point estimate and returning to the original assumption that there is a deterministic true solution. Although the solution to the inverse problem in the frequentist case is not a probability distribution, uncertainty quantification is still possible by studying confidence regions.

If we consider throwing a dice, then a frequentist would say that the probability of getting any given number from one to six is $\frac{1}{6}$. Let us then think that the person throwing the dice is a magician who was offering a hundred euros for rolling a six. A Bayesian could now take into account the prior information that there was probably a trick involved and lower the probability for getting a six. Unlike the frequentist concept, the Bayesian idea of probability is subjective and difficult to test. After observing the magician for a while the frequentist could also come to the same conclusion that the occurrence of six was indeed lower than $\frac{1}{6}$. The difference to the Bayesian paradigm is that in the frequentist framework the conclusion can only be drawn after observing a large number of throws and counting the relative frequency of six. Unfortunately, in practice such repetition is often impossible.

There are many similarities between classical regularization techniques and statistical methods. The function of prior information in the Bayesian scheme is to regularize the problem. Gaussian prior and noise distribution in the Bayesian approach produces the same point estimates as Tikhonov regularization methods, as we will see in the following sections. All the above methods have some advantages over the others and hence they can be seen to complement each other.

As mentioned before the regularized solutions with classical noise assumptions converge by design. Developing a similar comprehensive theory in the case of statistical noise assumption is important since stochastic models are often used in practice [19, 26, 36, 55, 68]. The main purpose of this thesis is to prove such convergence results when large noise is assumed. Paper [I] shows that convergence of continuous Tikhonov regularized solutions can be obtained in appropriate Sobolev scales when the white noise model is assumed. In paper [II] we prove convergence for more general regularized solutions in Banach spaces.

Another goal of this dissertation is to develop the theory of statistical inverse problems. As mentioned above Bayesian and frequentist methods have been used widely in practice but there are still many open questions in the area. The analysis of small noise limit in statistical inverse problems, also known as the theory of posterior consistency, has attracted a lot of interest in the last decade, see e.g. [1, 2, 20, 30, 41, 46, 45, 48, 52, 59, 66, 72, 77]. However, much remains to be done. In paper [II] we show some general convergence rates in the frequentist framework whereas paper [III] concentrates on the Bayesian and frequentist inverse problems with Gaussian noise and prior assumptions.

The results in papers [I] and [III] support the idea of regularization and Bayesian paradigm supporting and completing each other. Interpreting the Tikhonov regularized solution as a point estimate of the Bayesian inverse problem explains trivially some of the behaviour that is difficult to understand from a purely deterministic point of view. On the other hand regularization allows free choice of the regularization parameter. With a correct choice of the parameter we can show that the regularized solution converges in a Sobolev space with a smoothness index arbitrarily close to the smoothness of the true solution. We also prove the intuition that when larger noise is assumed, then stronger regularization is needed to guarantee the convergence to be true.

The rest of this dissertation is organised as follows. In Section 2 we introduce in more detail the background of regularization, frequentist and Bayesian approaches. We also define large noise from the statistical and deterministic perspectives and describe the *white noise paradox*. The first part of Section 3 considers the modifications we have to do for Tikhonov regularization to arrive at something useful with large noise assumptions. The convergence results of paper [I] are also described in detail. The rest of the section explains the variational regularization approach with convex regularization functional in Banach spaces. The deterministic and frequentist convergence results obtained in paper [II] are also presented. In Section 4 we explain the Bayesian paradigm along with the contraction and uncertainty quantification results studied in paper [III]. The implications of the results of papers [I-III] are discussed in Section 5 thus concluding the text.

2 Variational and stochastic inverse problems

2.1 Regularization methods

We are interested in the following continuous model

$$m = Au + \delta\varepsilon, \tag{2.1}$$

where the data m and the quantity of interest u are real-valued functions of d real variables. Above ε models the noise that is inevitable in practical measurements and

$\delta \in \mathbb{R}_+$ describes the noise amplitude. Here $\delta\varepsilon$ is just the product of $\delta > 0$ and ε. The forward operator $A : X \to Y$ is a bounded linear operator and X and Y are the model and measurement spaces respectively. A large class of practical measurements can be modelled by operators A arising from partial differential equations of mathematical physics. In ill-posed problems A does not have a continuous inverse.

In real life a physical measurement device produces a discrete data vector $\mathbf{m} \in \mathbb{R}^k$ instead of continuous function m. We model this by adding a device related linear operator P_k to (2.1):

$$\mathbf{m} := P_k(Au) + \delta P_k \varepsilon \tag{2.2}$$

and call (2.2) practical measurement model. As an example we can think of a case where u is an acoustic source and Au is acoustic pressure of the product acoustic wave. Then $P_k(Au) = \langle \phi_k, Au \rangle_{L^2(\mathbb{R}^3)}$ where ϕ_k can be thought to be the microphones used for measuring the data.

Usually nature does not offer a discretization of the unknown but we need a discrete representation of u to solve the problem in practice. Discretization of the unknown can be done using some computationally feasible approximation of the form $\mathbf{u} = T_n u \in \mathbb{R}^n$, for example Fourier series truncated to n terms. Then the practical inverse problem is

$$\textit{given a measurement } \boldsymbol{m}\textit{, estimate } \boldsymbol{u}. \tag{2.3}$$

The above problem has two independent discretizations since P_k is related to the measurement device and T_n to a (freely chosen) finite representation of the unknown.

In discrete case we can write the Tikhonov regularization in the form

$$\mathbf{u}_\alpha^T := \underset{\mathbf{u} \in \mathbb{R}^n}{\arg\min} \left\{ \|\mathbf{A}\mathbf{u} - \mathbf{m}\|_2^2 + \alpha \|\mathbf{L}\mathbf{u}\|_2^2 \right\} \tag{2.4}$$

where $\mathbf{A} = P_k A T_n$ is a $k \times n$ matrix approximation of the operator A. The first term on the right hand side of (2.4) is called the fidelity term and it ensures that the model is satisfied approximately. The regularization term $\|\mathbf{L}\mathbf{u}\|_2^2$ contains all our a priori knowledge of the solution. For example choosing $\mathbf{L} = \mathbf{I}$, identity matrix, we assume that the norm of the solution is not very large. If we assume $\mathbf{L} = \mathbf{I} + \mathbf{D}$, where $\mathbf{D}$ is a finite-difference first-order derivative matrix then our a priori assumption of the unknown is that u is continuously differentiable and u or its derivative is not very large in square norm. The regularization parameter $\alpha > 0$ can be used to tune the balance between the two requirements. Note that the minimization problem (2.4) is well defined with any choice of noise since we are summing up only a finite number of points.

The regularized solution $\mathbf{u}_\alpha^T$ can be written as

$$\mathbf{u}_\alpha^T = (\mathbf{A}^*\mathbf{A} + \alpha \mathbf{L}^*\mathbf{L})^{-1}\mathbf{A}^*\mathbf{m}.$$

One can then study the convergence of the approximated solution $\mathbf{u}_\alpha^T$ to the real solution $\mathbf{u}$.

Above the number k of data points is determined by the device while n can be chosen freely. Think for example electromagnetic measurements of brain activity. The unknown quantity is the current inside a patient's head that is modelled with a vector valued function $u = u(x)$, $x \in D \subset \mathbb{R}^3$. On the other hand, in numerical simulations the problem has to be discretized, which means that the continuous infinite dimensional model is approximated by a finite dimensional model. In this case the discretization is always done somewhat arbitrarily. Thus we face a problem of justification of the discretization. It is desirable that the reconstructions $\mathbf{u}_\alpha^T$ behave consistently when the measurement device is updated, that is, k is changed or when the computational grid is refined, meaning that n is increased. The latter may be required by a multigrid computational scheme or simply by a need of higher resolution in the reconstruction. By consistency we mean that the dependency of $\mathbf{u}_\alpha^T$ on k and n is stable, at least for large enough values.

If the discrete model is an orthogonal projection of the continuous model (2.1) to a finite dimensional subspace it guarantees that we can switch consistently between different discretizations. Hence a natural approach for ensuring consistency over k and n is to introduce a continuous version of (2.4). Under certain assumptions (including that the noise should be an L^2-function) the finite-dimensional problem (2.4) converges (in the sense of Γ-convergence [6]) as $n, k \to \infty$ to the following infinite-dimensional minimization problem in a Sobolev space H^r:

$$\arg\min_{u \in H^r} \left\{ \|m - Au\|_{L^2}^2 + \alpha \|u\|_{H^r}^2 \right\}. \tag{2.5}$$

The case $\mathbf{L} = \mathbf{I}$ in (2.4) corresponds to $r = 0$ and $\mathbf{L} = \mathbf{I} + \mathbf{D}$ corresponds, roughly to $r = 1$. Note that the above minimization problem (2.5) is well defined only when the noise ε is an L^2 function. The regularized solution of the continuous problem (2.5) can be written as

$$u_\alpha^T = (A^*A + \alpha(I - \Delta)^r)^{-1} A^* m. \tag{2.6}$$

In the Tikhonov regularization method above we assumed that the regularization term is a squared norm. Such regularization guarantees a noise robust solution but it also forces some smoothness to the regularized solution. Think about an inverse problem of recovering a sharp approximation from a blurred image. Using a squared norm type regularization term tends to promote smooth reconstructions. Instead we need a regularization term that allows quick jumps in the solution. One popular example of such an edge preserving regularization term is total variation, also covered in our work [II], which favours piecewise smooth functions that have rapidly changing values only in a set of small measure [11, 29, 64].

We can generalize the continuous Tikhonov regularization by looking for an approximated solution that is the minimizer of the square residual in the norm of Hilbert space Y and a convex regularization functional $R(u)$. That is, we are interested in solving a minimization problem

$$u_\alpha^R = \arg\min_{u\in X} \left\{ \frac{1}{2}\|Au - m\|_Y^2 + \alpha R(u) \right\}, \tag{2.7}$$

with a convex regularization functional $R : X \to \mathbb{R} \cup \{\infty\}$. Here it is enough to assume that X is a separable Banach space.

As mentioned before regularization with a convex regularization functional lets us model much wider range of a priori knowledge of the unknown than the quadratic regularization. In particular it includes one homogeneous regularization popularized by total variation and sparsity methods see e.g [8, 21, 58]. On the other hand convexity restriction offers us possibility to use the powerful machinery of convex analysis.

To guarantee the existence of the minimizer u_α^R we need the following assumptions on R in addition to the convexity:

(R1) the functional R is lower semicontinuous in some topology τ on X,

(R2) the sub-level sets $M_\rho = \{u \in X \mid R(u) \le \rho\}$ are compact in the topology τ on X and

Since we have a general convex regularization functional R instead of a squared norm in Hilbert space we do not get a solution in a similar formula as in (2.6). Instead we are looking for a minimizer u_α^R that fulfills the optimality condition

$$A^*(Au_\alpha^R - m) + \alpha\xi_\alpha = 0, \tag{2.8}$$

with some $\xi_\alpha \in \partial R(u_\alpha^R)$. Here

$$\partial R(u) = \{\xi \in X^* \mid R(u) - R(v) \le \langle \xi, u - v\rangle_{X^*\times X} \text{ for all } v \in X\}$$

stands for the subdifferential. Subdifferential generalizes the derivative to convex functions which are not differentiable. Note that subdifferential is not necessarily single valued.

We are interested in the error estimates between u_α^R and a solution u^* minimizing R among all possible solutions of $Au = m$. By modifying (2.8) and then taking a duality product with $u_\alpha^R - u^*$ we arrive to

$$\|A(u_\alpha - u^*)\|_Y^2 + \alpha D_R^{\xi_\alpha,\xi^*}(u_\alpha, u^*) \le \langle \delta A^*\varepsilon - \alpha\xi^*, u_\alpha - u^*\rangle_{X^*\times X} \tag{2.9}$$

where $D_R^{\xi_\alpha,\xi^*}(u_\alpha, u^*)$ is the symmetric Bregman distance defined by

$$D_R^{\xi_u,\xi_v}(u, v) = \langle \xi_u - \xi_v, u - v\rangle_{X^*\times X}$$

for all $\xi_v \in \partial R(v)$, $\xi_u \in \partial R(u)$ and $u, v \in X$. The Bregman distance is routinely used for error estimation in regularization, see e.g. [3, 7, 9, 12, 31, 37, 43, 53, 61, 62]. In the quadratic case $R(u) = \|u\|_X^2$ where X is a Hilbert space Bregman distance coincides with the squared norm

$$D_R^{\xi_u,\xi_v}(u, v) = \|u - v\|_X^2$$

and hence it is a natural generalization for the classical error estimation.

The nice case leading directly to estimates is to assume that the unknown fulfills source condition $\xi^* = K^*w^* \in X^*$ with some $w^* \in Y$ and classical noise, that is, $\varepsilon \in Y$. Then Young's inequality gives us an estimate

$$D_R^{\xi_\alpha,\xi^*}(u_\alpha, u^*) \leq \frac{1}{2\alpha}\|\delta\varepsilon - \alpha w^*\|_Y^2$$

and we can find optimal regularization strategy by minimizing $\alpha = \alpha(\delta)$.

2.2 Stochastic inverse problems

Another approach to finding a noise robust solution for an inverse problem is to study it from Bayesian point of view [17, 27, 42, 50, 69]. This means that instead of the deterministic problem (2.1) we are interested in the model

$$M_\delta = AU + \delta\mathcal{E} \qquad (2.10)$$

where the measurement $M_\delta = M_\delta(\omega)$, the unknown $U = U(\omega)$ and the noise $\mathcal{E} = \mathcal{E}(\omega)$ are modelled as random variables. Here $\omega \in \Omega$ is an element of a complete probability space $(\Omega, \Sigma, \mathbb{P})$. The philosophical reason why we model also U as a random variable is that even though the unknown quantity is assumed to be deterministic we have only incomplete information about it. That is, the randomness of U is not thought to be a property of the unknown but of the observer [5, 13].

All information available about the unknown before performing the measurements is included in a priori distribution which is independent of the measurement. One of the core difficulties of Bayesian inverse problems is to encode the known properties of U to a probability distribution.

As in the deterministic case to study the model (2.10) in practice we need to discretize it. We can do this in a similar way as in the previous section. Assume now that the measurement $\mathbf{M}$ and the noise $\mathbf{E}$ take values in $\mathbb{R}^k$ and the unknown $\mathbf{U}$ in $\mathbb{R}^n$. To solve the inverse problem

$$\textit{given a realization of } \boldsymbol{M}\textit{, estimate } \boldsymbol{U} \tag{2.11}$$

we have to express available a priori information of $\mathbf{U}$ in the form of a probability density π_{pr} in an n-dimensional subspace. We denote the densities of $\mathbf{M}$ and $\mathbf{E}$ by $\pi_{\mathbf{M}}$ and $\pi_{\mathbf{E}}$, respectively. The solution of the Bayesian inverse problem after performing the measurements is the posterior distribution of the unknown random variable. Computational exploration of the finite-dimensional posterior distribution yields useful estimates of the quantity of interest and enables uncertainty quantification. Furthermore, analytic results about the continuous model can then be restricted to a given resolution in a discretization-invariant way.

The Bayesian inversion theory is based on the Bayes formula. Given a realization of the discrete measurement the posterior density for $\mathbf{U}$ taking values in the n-dimensional subspace is given by the Bayes formula

$$\pi(\mathbf{u}\,|\,\mathbf{m}) = \frac{\pi_{pr}(\mathbf{u})\pi_{\mathbf{E}}(\mathbf{m}\mid\mathbf{u})}{\pi_{\mathbf{M}}(\mathbf{m})} \quad \mathbf{u}\in\mathbb{R}^n,\ \mathbf{m}\in\mathbb{R}^k. \tag{2.12}$$

An approximated solution for the inverse problem is often given as a point estimate for (2.12). The maximum a posteriori (MAP) estimator $T_\delta^{MAP}:\mathbb{R}^k\to\mathbb{R}^n$ is defined by

$$T_\delta^{MAP}(\mathbf{M}(\omega)) := \arg\max_{\mathbf{u}\in\mathbb{R}^n} \pi(\mathbf{u}\,|\,\mathbf{M}(\omega)). \tag{2.13}$$

Note that the MAP estimate depends on ω through the realization of the noise $\mathbf{E}(\omega)$ and unknown $\mathbf{U}(\omega)$. Another often used point estimate is conditional mean (CM) estimate defined by

$$T_\delta^{CM}(\mathbf{M}(\omega)) = \mathbb{E}(\mathbf{U}|\mathcal{M})(\omega) \quad \text{a.s.} \tag{2.14}$$

where $\mathcal{M}$ is the σ-algebra generated by $\mathbf{M}$. If we assume white Gaussian noise (see Section 2.3 for the exact definition) and Gaussian prior distribution the MAP and CM estimates coincide a.s.

Let us denote the covariance matrix of $\mathbf{U}$ by $\mathbf{C_U}$. In Gaussian case solving the maximization problem (2.13) with a fixed realization of noise and unknown corresponds to solving the minimization problem

$$\mathbf{u}_\delta^B = \arg\min_{\mathbf{u}\in\mathbb{R}^n}\left\{\frac{1}{2\delta^2}\|\mathbf{A}\mathbf{u}-\mathbf{m}\|_2^2 + \frac{1}{2}\|\mathbf{C_U}^{-1/2}\mathbf{u}\|_2^2\right\}. \tag{2.15}$$

That is, in Gaussian case the MAP estimate coincides with Tikhonov regularized solution where $\alpha=\delta^2$ and $\mathbf{L}=\mathbf{C_U}^{-1/2}$.

In infinite dimensional Bayesian inverse problems the problem arises from the fact that there is no continuous equivalent to Bayes formula. The posterior distribution can be formulated using the Radon–Nikodym derivative but it is usually challenging to calculate explicitly. If we assume Gaussian prior and noise the posterior distribution is also Gaussian and the mean and covariance can be calculated explicitly.

As before we are interested in the convergence properties of the approximated solution. Since the point estimate $U_\delta^B(\omega)$ depends on the realization of the prior and noise we are interested in the following convergence

$$\mathbb{E}\|U_\delta^B(\omega) - U(\omega)\|_{H^\varsigma(N)} \to 0, \quad \text{as } \delta \to 0,$$

where the expectation $\mathbb{E}$ is taken with respect to U and $\mathcal{E}$. Combined with the convergence of the covariance operator the convergence of the CM estimate guarantees the contraction of the posterior distribution.

Stochastic inverse problems can also be studied from frequentist point of view. Then one is interested in a model

$$M_\delta^\dagger(\omega) = A(u^\dagger) + \delta\mathcal{E}(\omega) \tag{2.16}$$

where the data $M_\delta^\dagger$ is generated by a true solution $u^\dagger$ instead of random draw $U(\omega)$ from the prior distribution. This means that in (2.16) all the randomness of the $M_\delta^\dagger$ comes from the randomness of the noise $\mathcal{E}$. The main interest is then on the contraction of the posterior distribution around the true solution $u^\dagger$ as the noise goes to zero.

2.3 Large noise

In classical regularization theory the noise term is assumed to be deterministic and small. In such a case one has a norm estimate of the noise and can design regularization strategies such that $u_{\alpha(\delta)} \to u$ as $\delta \to 0$. This approach has been studied in depth and the literature on the topic is extensive see e.g [10, 25, 33, 44, 54, 56, 57, 74].

We, however, are interested in stochastic modelling of noise which includes the classical small noise but allows also wider modelling of ε. Generally large noise means that the norm of the data perturbation introduced by the noise is not small or it can even be unbounded in the image space of the forward operator. Statistical modeling of noise in the inverse problems started in the early papers of [28, 27, 70, 73].

There has been several papers tackling the problem of large noise in the settings of regularization methods. One way is to assume that the noise is potentially large in the image space of the forward operator but still an element of that space. This idea of weakly bounded noise was introduced in papers [23, 24, 22]. Such a relaxed assumption of noise covers small low frequency noise and large high frequency noise. However, even though $\delta\varepsilon$ tends to zero in weak sense as $\delta \to 0$ and ε is a realization of white noise, this type of noise lies outside the definition of the weakly bounded noise since white noise takes values in image space Y only with probability zero as we will see below.

Our interest in large noise is motivated by stochastic modelling of noise and especially by the white noise model. One reason we are interested in white Gaussian noise is that the central limit theorem indicates that the summation of many random processes will tend to have Gaussian distribution. Any Gaussian noise can then be whitened rendering white noise model. Next we will give definitions for the discrete and continuous white noise and describe the white noise paradox arising from the infinite L^2-norm of the natural limit of white Gaussian noise in $\mathbb{R}^k$ when $k \to \infty$.

We model the k-dimensional noise $P_k\varepsilon$ as a vector $\mathbf{e} \in \mathbb{R}^k$. Here $\mathbf{e}$ is a realization of a $\mathbb{R}^k$-valued Gaussian random variable $\mathbf{E}$ having mean zero and unit variance: $\mathbf{E} \sim N(0, I)$. In terms of a probability density function we have

$$\pi_{\mathbf{E}}(\mathbf{e}) = c \exp\Big(-\frac{1}{2}\|\mathbf{e}\|_2^2 \Big). \tag{2.17}$$

The appearance of $\|\cdot\|_2$ in (2.17) is the reason why square norm is used in the data fidelity term $\|\mathbf{A}\mathbf{u}-\mathbf{m}\|_2^2$. The above noise model is appropriate for example in photon counting under high radiation intensity, see e.g. [49, 68].

Let N be a closed d-dimensional manifold. We assume N to be closed to simplify the settings so that we do not have to study boundary value problems. Continuous white noise $\mathcal{E}$ can be considered as a measurable map $\mathcal{E} : \Omega \to \mathcal{D}'(N)$ where Ω is the probability space. Then normalized white noise is a random generalized function $\mathcal{E}(x, \omega)$ on N for which the pairings $\langle \mathcal{E}, \phi \rangle_{\mathcal{D}' \times \mathcal{D}}$ are Gaussian random variables for all test functions $\phi \in \mathcal{D} = C^\infty(N)$, $\mathbb{E}\mathcal{E} = 0$, and

$$\mathbb{E}\Big(\langle \mathcal{E}, \phi \rangle_{\mathcal{D}' \times \mathcal{D}} \langle \mathcal{E}, \psi \rangle_{\mathcal{D}' \times \mathcal{D}} \Big) = \langle I\phi, \psi \rangle_{\mathcal{D}' \times \mathcal{D}} = \int_N \phi(x)\psi(x) dV_g(x) \tag{2.18}$$

for $\phi, \psi \in \mathcal{D}$. Above $dV_g(x)$ is the volume form. Non-rigorously, this is often written as $\mathbb{E}\big(\mathcal{E}(x)\mathcal{E}(y)\big) = \delta_y(x)$. We will denote this by $\mathcal{E} \sim N(0, I)$. A realization of $\mathcal{E}$ is the generalized function $\varepsilon = \mathcal{E}(\,\cdot\,, \omega_0)$ on N with a fixed $\omega_0 \in \Omega$.

The probability density function of white noise $\mathcal{E}$ is often *formally* written in the form

$$\pi_{\mathcal{E}}(\varepsilon) \underset{formally}{=} c \exp\Big(-\frac{1}{2}\|\varepsilon\|_{L^2(N)}^2 \Big). \tag{2.19}$$

Note that even though (2.17) is well defined with any $k \in \mathbb{R}$ the limit of the norm $\|\mathbf{e}_k\|_2^2$ is infinite when $k \to \infty$. Hence the above density function (2.19) is not well defined and can be thought only as a formal limit to (2.17). We will next illustrate the fact that the realizations of white Gaussian noise are almost surely not in $L^2(N)$ by an example in a d-dimensional torus $\mathbb{T}^d$.

Let $\mathcal{E}$ be normalized white Gaussian noise defined on the d-dimensional torus $\mathbb{T}^d = (\mathbb{R}/(2\pi\mathbb{Z}))^d$. The Fourier coefficients of $\mathcal{E}$ are normally distributed with variance one, that is, $\langle \mathcal{E}, e_{\vec{\ell}} \rangle \sim N(0,1)$, where $e_{\vec{\ell}}(x) = e^{i\vec{\ell}\cdot x}$ and $\vec{\ell} \in \mathbb{Z}^d$. Then

$$\mathbb{E}\|\mathcal{E}\|^2_{L^2(\mathbb{T}^d)} = \sum_{\vec{\ell}\in\mathbb{Z}^d} \mathbb{E}|\langle \mathcal{E}, e_{\vec{\ell}} \rangle|^2 = \sum_{\vec{\ell}\in\mathbb{Z}^d} 1 = \infty.$$

This implies that realizations of $\mathcal{E}$ are in $L^2(\mathbb{T}^d)$ with probability zero. However, when $s > d/2$

$$\mathbb{E}\|\mathcal{E}\|^2_{H^{-s}(\mathbb{T}^d)} = \sum_{\vec{k}\in\mathbb{Z}^d} (1 + |\vec{\ell}|^2)^{-s} \mathbb{E}|\langle \mathcal{E}, e_\ell \rangle|^2 < \infty \tag{2.20}$$

and hence $\mathcal{E}$ takes values in $H^{-s}(\mathbb{T}^d)$ almost surely (that is, with probability one).

On the other hand [63, Theorem 2] implies that if $\|\mathcal{E}\|^2_{H^{-s}(\mathbb{T}^d)} < \infty$ almost surely then $\mathbb{E}\|\mathcal{E}\|^2_{H^{-s}(\mathbb{T}^d)} < \infty$ which yields $s > d/2$. This concludes that the realizations of white noise $\mathcal{E}$ are almost surely in the space $H^{-s}(\mathbb{T}^d)$ if and only if $s > d/2$. In particular for $s \leq d/2$ the function $x \mapsto \mathcal{E}(x,\omega)$ is in $H^{-s}(\mathbb{T}^d)$ only when $\omega \in \Omega_0 \subset \Omega$ where $\mathbb{P}(\Omega_0) = 0$.

Motivated by the above stochastic modelling of white Gaussian noise, we assume in the Sobolev space regularization framework in the paper [I] that $\varepsilon \in H^{-s}(N)$ with some $s > d/2$. In more general regularization settings in Banach spaces with a general convex regularization functional R [II] we assume that the noise takes values in a Banach space Z^*. Here Z^* is a part of the Gelfand triple (Z, Y, Z^*) where $Z \subset Y$ is a dense subspace with Banach structure and the dual pairing of Z and Z^* is compatible with the inner product of a Hilbert space Y, i.e., by identifying $Y = Y^*$ we have

$$\langle u, v \rangle_{Z\times Z^*} = \langle u, v \rangle_Y$$

whenever $u \in Z \subset Y$ and $v \in Y = Y^* \subset Z^*$. Relating to the above mentioned Sobolev scales we can take as an example Gelfand triple (Z, Y, Z^*) where $Z = H^s(N)$, $Y = L^2(N)$, and $Z^* = H^{-s}(N)$.

2.4 Pseudodifferential operators and hypoellipticity

In papers [I] and [III] we study the measurement model (2.1) where the forward operator A is assumed to be an elliptic or hypoelliptic pseudodifferential operator. Pseudodifferential operators are a generalization of differential operators written in a form of Fourier integral operators. We can define the class of pseudodifferential operators as follows.

Let $m \in \mathbb{R}$. The symbol class $S^m(\mathbb{R}^d, \mathbb{R}^d)$ consists of such $a(x,\xi) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ that for all multi-indices α and β and any compact set $K \subset \mathbb{R}^d$ there exists a constant $C_{\alpha,\beta,K} > 0$ for which

$$|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)| \leq C_{\alpha,\beta,K}(1+|\xi|)^{m-|\alpha|}, \qquad \xi \in \mathbb{R}^d,\ x \in K.$$

A bounded linear operator $A : \mathcal{D}'(\mathbb{R}^d) \to \mathcal{D}'(\mathbb{R}^d)$ is called a pseudodifferential operator of order m if there is a symbol $a \in S^m(\mathbb{R}^d \times \mathbb{R}^d)$ such that for $u \in C^\infty(\mathbb{R}^d)$ we have

$$Au(x) = \int_{\mathbb{R}^d} e^{i(x-y)\cdot\xi} a(x,\xi) u(y) dy d\xi.$$

As an example we can think of a forward operator A that is defined by

$$Au(x) = \int_N \mathcal{A}(x,z) u(z) dz$$

where $\mathcal{A} \in C^\infty\big((\mathbb{R}^d \times \mathbb{R}^d)\setminus \mathrm{diag}(\mathbb{R}^d)\big)$ and in an open neighbourhood $V \Subset \mathbb{R}^d \times \mathbb{R}^d$ of the $\mathrm{diag}(\mathbb{R}^d) = \{(x,x);\ x \in \mathbb{R}^d\}$, we have

$$\mathcal{A}(x,z) = \frac{b(x,z)}{d_g(x,z)^p}, \quad (x,z) \in V$$

where d_g is a distance function, $p < d$, $b \in C^\infty(V)$ and $b(x,x) \neq 0$. In this case $\mathcal{A}$ is a pseudodifferential operator of order $-d+p<0$.

A pseudodifferential operator A is called elliptic if its principal symbol $a_m(x,\xi)$ satisfies

$$a_m(x,\xi) \neq 0 \qquad \text{for} \quad (x,\xi) \in \mathbb{R}^d \times (\mathbb{R}^d \setminus 0).$$

In paper [III] we are interested in a more general class of hypoelliptic operators. Let $t, t_0 \in \mathbb{R}$. We define symbol class $HS^{-t,-t_0}(\mathbb{R}^d, \mathbb{R}^d)$ to consist of $a(x,\xi) \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ for which

1. For an arbitrary compact set $K \subset \mathbb{R}^d$ we can find such positive constants R, c_1 and c_2 that

$$c_1(1+|\xi|)^{-t_0} \le |a(x,\xi)| \le c_2(1+|\xi|)^{-t}, \qquad |\xi| \ge R,\ x \in K.$$

2. For any compact set $K \subset \mathbb{R}^d$ there exist constants R and $C_{\alpha,\beta,K}$ such that for all multi-indices α and β

$$|\partial_\xi^\alpha \partial_x^\beta a(x,\xi)| \le C_{\alpha,\beta,K}|a(x,\xi)|(1+|\xi|)^{-|\alpha|}, \qquad |\xi| \ge R,\ x \in K.$$

The pseudodifferential operators with symbol $a(x,\xi) \in HS^{-t,-t_0}(V \times \mathbb{R}^d)$ are called hypoelliptic. Note that a hypoelliptic operator A is elliptic if $t = t_0$. One example of a hypoelliptic operator that is not elliptic is the heat operator

$$Pu(x,t) = \partial_t u - k\Delta_x u, \quad (x,t) \in \mathbb{R}^d \times \mathbb{R}.$$

In the Bayesian approach the covariance operators $C_\mathcal{E} = I$ and C_U are assumed to be elliptic. If we assume that also the forward operator A is elliptic then $A \simeq C_U^\gamma$ with some $\gamma \in \mathbb{R}$. Here $\simeq$ is used loosely to indicate two operators which induce equivalent norms. By allowing A to be hypoelliptic we can study a much wider range of problems where the model and the prior do not have to be as strongly connected as in the elliptic case.

3 Regularization results

3.1 Modification of Tikhonov regularization for large noise

If we assume the noise in (2.1) to be large then the fidelity term in the minimization problem (2.5) is not well defined. To overcome this problem in paper [I] we have modified the Tikhonov regularization to arrive at something useful for large noise $\varepsilon \in H^{-s}(N)$, $s > d/2$.

If the noise term is an $L^2(N)$ function we can write

$$\|m - Au\|^2_{L^2(N)} = \|Au\|^2_{L^2(N)} - 2\langle m, Au\rangle_{L^2(N)} + \|m\|^2_{L^2(N)}.$$

Omitting the constant term $\|m\|^2_{L^2(N)}$ in (2.5) leads to a definition

$$u_\alpha^T = \underset{u \in H^r(N)}{\arg\min} \left\{ \|Au\|^2_{L^2(N)} - 2\langle m, Au\rangle + \alpha\|u\|^2_{H^r(N)} \right\}, \tag{3.1}$$

where we can interpret $\langle m, Au\rangle$ as a suitable duality pairing instead of $L^2(N)$ inner product. When A is a pseudodifferential operator of order $-t \le r - s$, we can define $\langle m, Au\rangle = \langle m, Au\rangle_{H^{-s}(N)\times H^s(N)}$. Then the regularized solution u_α^T is well defined even when $\varepsilon \notin L^2(N)$ as long as the forward operator A is smoothing enough. Note that when $\varepsilon \in L^2(N)$ minimization problems (2.5) and (3.1) have the same solution.

The regularized solution of the modified problem (3.1) can be written in the form

$$u_\alpha^T = (A^*A + \alpha(I-\Delta)^r)^{-1}A^*m.$$

We have chosen the regularization parameter to be a function of the noise amplitude: $\alpha(\delta) = \alpha_0\delta^\kappa$, where $\alpha_0 > 0$ is a constant and $\kappa > 0$.

Using the microlocal analysis and Shubin calculus we can prove the following convergence theorem [I]. For general theory see [40, 67]. Microlocal analysis has been used successfully in study of inverse problems see for example [32].

Theorem 1 *Let N be a d-dimensional closed manifold and $u \in H^r(N)$ with $r \ge 0$. Here $\|u\|_{H^r(N)} := \|(I-\Delta)^{r/2}u\|_{L^2(N)}$. Let $\varepsilon \in H^{-s}(N)$ with some $s > d/2$ and consider the measurement*

$$m_\delta = Au + \delta\varepsilon, \tag{3.2}$$

where A, is an elliptic pseudodifferential operator of order $-t$ on the manifold N with $t > \max\{0, s-r\}$ and $\delta \in \mathbb{R}_+$. Assume that $A : L^2(N) \to L^2(N)$ is injective. The regularization parameter is chosen to be $\alpha(\delta) = \alpha_0\delta^\kappa$, where $\alpha_0 > 0$ is a constant and $\kappa > 0$.

Take $\zeta \le 2(t+r)/\kappa - s - t$. Then the following convergence takes place in $H^{s_1}(N)$ norm:

$$\lim_{\delta\to 0} u^T_{\alpha(\delta)} = u.$$

Furthermore, we have the following estimates for the speed of convergence:

(i) If $\zeta \le -s-t$ then

$$\|u_\alpha^T - u\|_{H^{s_1}} \le C\max\{\delta^{\frac{\kappa(r-\eta)}{2(t+r)}}, \delta\}. \tag{3.3}$$

(ii) If $-s-t \le \zeta < 2(t+r)/\kappa - s - t$ then

$$\|u_\alpha^T - u\|_{H^{s_1}} \le C\max\{\delta^{\frac{\kappa(r-\eta)}{2(t+r)}}, \delta^{1-\frac{\kappa(s+t+\zeta)}{2(t+r)}}\}. \tag{3.4}$$

Above we have $\eta = \max\{\zeta, -r - 2t\}$.

From the above Theorem 1 we see that when $\kappa \leq 1$ the approximated solution $u_\alpha^T \in H^r(N)$ converges to the real solution $u \in H^r(N)$ in space $H^{r-\epsilon}(N)$, with arbitrary small $\epsilon > 0$. In comparison, in the classical regularization theory one only needs to assume $\kappa < 2$ for convergence. Looking at the formula (2.5) we see that when $\varepsilon \in L^2(N)$ the fidelity term can be written $\|m - Au\|^2_{L^2(N)} = \delta^2 \|\varepsilon\|^2_{L^2(N)}$. Then the regularization term $\alpha \|u\|_{H^r(N)}$ has the same asymptotic behaviour when $\kappa = 2$. Since the problem is ill-posed regularization is needed also with small δ and hence one needs to assume $\kappa < 2$ to get a robust solution. When large noise is assumed it is natural that stronger regularization, that is, smaller κ is needed to guarantee the convergence of the regularized solution u_α^T. We also notice that the smoother the forward operator A is the worse convergence rates we get.

We can also offer counter examples showing that with the wrong choice of κ the regularized solution u_α^T diverges when $\delta \to 0$. Such behaviour can already be seen in the discrete settings when the discretization is chosen to be fine enough. This underlines the importance of understanding the connection between a discrete model and its infinite-dimensional limit model. Lack of convergence in the continuous inverse problem can lead to slow algorithms for the practical problem.

Since the operator A does not have a continuous inverse operator $L^2 \to L^2$, the condition number of the matrix approximation $\mathbf{A}$ of the operator A grows when the discretization is refined. This is the very reason why regularization is needed in the (numerical) solutions of the inverse problems. We can demonstrate this problem by an example.

Consider the inverse problem (2.1) in two-dimensional torus on $\mathbb{T}^2$. We assume noise to be a realization of white Gaussian noise, that is, $\varepsilon \in H^{-s}(\mathbb{T}^2)$ with $s > 1$. The forward operator A is assumed to be an elliptic operator, smoothing of order 2,

$$(Au)(x) = \mathcal{F}^{-1}\big((1 + |n|^2)^{-1}(\mathcal{F}u)(n)\big)(x).$$

Solving u from $Au(x) = m(x)$ corresponds to the solution of the ordinary differential equation $(1 - \partial_x^2)m(x) = u(x)$ so A can be thought e.g. as a blurring operator.

The unknown is an H^1 function shown in Figure 1.

The approximated solution to the problem is

$$u_\delta^T = (A^*A + \delta^2(I - \Delta))^{-1} A^* m_\delta.$$

Note that above we have $\alpha = \delta^2$, that is, κ is chosen to be too large. Theorem 1 guarantees convergence

$$\lim_{\delta \to 0} \|u_\alpha^T - u\|_{H^\zeta} = 0$$

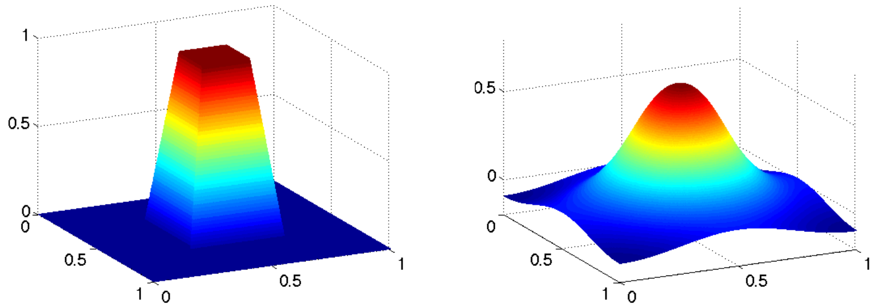

Figure 1: On the left the original piecewise linear function $u \in H^1(\mathbb{T}^2)$. On the right side the noiseless data $m = Au$.

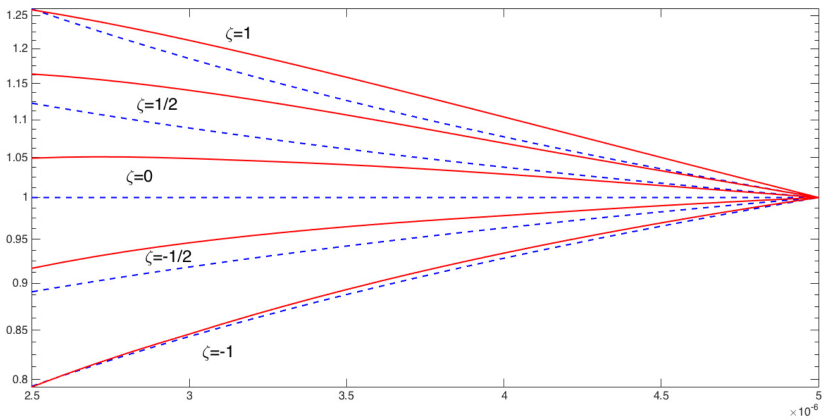

Figure 2: Normalized errors $c(\zeta)\|u_\alpha^T - u\|_{H^\zeta(\mathbb{T}^2)}$ in logarithmic scale with different values of ζ. The numerically solved errors $c(\zeta)\|u_\alpha^T - u\|_{H^\zeta(\mathbb{T}^2)}$, for the example u given in Figure 1, are plotted with solid lines and the bounds (3.4) given in Theorem 1 are plotted with dashed lines.

when $\zeta < -\tau < 0$. This behaviour can be seen even in numerical simulations when the discretization is fine enough. In Figure 2 we have compared the expected convergence rates given in formula (3.4) in Theorem 1 to the computational convergence rates. We see that for the test case presented in Figure 1 the convergence $u_\alpha^T \to u$ in different Sobolev spaces follows well the convergence predicted by Theorem 1.

3.2 Variational regularization

We will now proceed to study regularization with a more general regularization functional R in a separable Banach space X. For our setting of the noise let (Z, Y, Z^*) be a Gelfand triple such that $Z \subset Y$ is a dense subspace with Banach structure and the dual pairing of Z and Z^* is compatible with the inner product of Y, i.e., by identifying $Y = Y^*$ we have

$$\langle u, v\rangle_{Z\times Z^*} = \langle u, v\rangle_Y$$

whenever $u \in Z \subset Y$ and $v \in Y = Y^* \subset Z^*$. We then assume that $\varepsilon \in Z^*$. The key assumption we make is that $A : X \to Z$ is continuous. It directly follows that A^* has a continuous extension $A^* : Z^* \to X^*$. It is crucial that due to the continuous extension property $A^*\varepsilon$ is bounded in X^*.

As in the Tikhonov case we we need to modify the fidelity term to get a well defined estimate in case of large noise. We are interested in solving a minimization problem

$$u_\alpha^R = \arg\min_{u\in X} \Big\{ \|Au\|_Y^2 - 2\langle m, Au\rangle + \alpha R(u) \Big\}, \tag{3.5}$$

with a convex regularization functional $R : X \to \mathbb{R} \cup \{\infty\}$.

Now the question is: when does the minimization problem (3.5) have a unique minimizer? To guarantee the existence of the minimizer in case of large noise we need one more assumption in addition to (R1) and (R2) given in Section 2.1:

(R3) the convex conjugate R^* is finite on a ball in X^* centered at zero.

Above the convex conjugate $R^* : X^* \to \mathbb{R} \cup \{\infty\}$ is defined by

$$R^*(q) = \sup_{u\in X} \left(\langle q, u\rangle_{X^*\times X} - R(u) \right).$$

The major difficulty in the case of large noise is that there is no natural lower bound for (3.5). In the case of bounded noise we immediately see that the problem is bounded below by $-\frac{1}{2}\|m\|_Y^2 + \alpha R(u_0)$, with u_0 being a minimizer of R. However, this problem can be overcome by suitable approximation of the noise together with

(R3) and the lower bound then guarantees the existence of the unique minimizer, see [II].

Let us rewrite (2.9) in form

$$\|A(u_\alpha^R - u^*)\|_Y^2 + \alpha D_R^{\xi_\alpha,\xi^*}(u_\alpha^R, u^*) \leq \langle \delta\eta - \alpha\xi^*, u_\alpha^R - u^* \rangle_{X^* \times X} \tag{3.6}$$

where $\eta = A^*\varepsilon$. The above implies that the assumption $\varepsilon \in Y$ and the source condition for the unknown play a similar role in the classical regularization and a violation of either of them leads to similar problems in the analysis. This means that technically η not in the range of A^* is equally difficult as ξ^* not in the range of A^*. Here the range is defined as A^*Y and not A^* on a larger space including the noise.

The case of ξ^* not fulfilling the source condition is reasonably well understood, at least in the case of strictly convex functionals R, see [65]. The idea is to use a so-called approximate source condition, quantifying how well ξ^* can be approximated by elements in the range of A^*. Since ξ^* needs to be in the closure of the range, there exists a sequence w_n^* with $A^*w_n^* \to \xi^*$. On the other hand it is not in the range, hence $\|w_n^*\|$ necessarily diverges. Thus one can measure how well ξ^* can be approximated by elements A^*w^* with a given upper bound on $\|w^*\|$. The best estimates are then obtained by balancing errors containing the approximation of ξ^* and $\|w^*\|$.

In the case of no strict source condition and unbounded noise one can approximate ξ^* and η with separate elements A^*w_1 and A^*w_2 respectively. Then the right hand side of (3.6) can be written in the form

$$\begin{aligned}&\langle \delta\eta - \alpha\xi^*, u_\alpha^R - u^* \rangle_{X^* \times X} = \\ &\qquad \langle \delta(\eta - A^*w_2) - \alpha(\xi^* - A^*w_1), u_\alpha^R - u^* \rangle_{X^* \times X} + \langle \delta w_2 - \alpha w_1, A(u_\alpha^R - u^*) \rangle_Y,\end{aligned}$$

where $w_1, w_2 \in Y$. The second term on the right hand side can now be estimated using Young's inequality as in the case of small noise and source condition. For the first term it is natural to apply the generalized Young's inequality

$$\begin{aligned}\left\langle \xi^* - A^*w_1, u_\alpha^R - u^* \right\rangle_{X^* \times X} &= \zeta \left\langle \frac{\xi^* - A^*w_1}{\zeta}, u_\alpha^R - u^* \right\rangle_{X^* \times X} \\ &\leq \zeta R\left(u_\alpha^R - u^*\right) + \zeta R^*\left(\frac{\xi^* - A^*w_1}{\zeta}\right),\end{aligned}$$

which we shall employ further with appropriately chosen $\zeta > 0$. We observe that in proceeding as above we are left with two terms in dependence on w_1, namely $\frac{\alpha^2}{2}\|w_1\|^2$ and $\alpha\zeta R^*(\frac{A^*w_1 - \xi^*}{\zeta})$. This motivates our approach to the approximate source conditions to be detailed in the following.

3.3 Approximated source condition

As mentioned before the case where the unknown does not fulfill the strict source condition is reasonably well understood and has been tackled by the concept of distance functions and approximate source conditions [35, 38, 39, 65]. The standard concept of approximate source condition is to consider the case $R(u) = \|u\|_X^r$ for some power $r > 1$ (cf. [65]). The approximated source condition is then defined via distance function below

$$d_\rho(\xi^*) := \inf_{w \in Y} \{\|\xi^* - A^* w\|_{X^*} \mid \|w\|_Y \leq \rho\} \tag{3.7}$$

and one is interested in the asymptotics of (3.7) as $\rho \to \infty$. Note that in the case of a fulfilled source condition $d_\rho(\xi^*) = 0$ for ρ sufficiently large, while in the really approximate case $d_\rho(\xi^*)$ decays to zero at a finite rate. Hence, the speed of decay of $d_\rho(\xi^*)$ is a natural measure to quantify the approximateness of the source condition. Unfortunately the existing theory employing the approximate source conditions or the even more implicit variational inequalities only works for the special norm-type functionals above (cf. [65]) and in addition uses some moduli of strict convexity of the norms. This of course excludes the most interesting cases of one-homogeneous regularizations such as sparsity and total variation. Hence we propose to consider a more general formulation based on convex duality.

Due to the analogous role of η and ξ^* it is natural to use the same paradigm for approximating both of them. As we have seen above we want to approximate elements $\xi \in X^*$ by $A^* w$ with $w \in Y$. More precisely, we are interested in minimal values of the functional

$$E_{\alpha,\zeta}(w; \xi) = \zeta R^* \left(\frac{\xi - A^* w}{\zeta} \right) + \frac{\alpha}{2} \|w\|_Y^2,$$

which we shall denote as

$$e_{\alpha,\zeta}(\xi) = \inf_{w \in Y} E_{\alpha,\zeta}(w; \xi). \tag{3.8}$$

In the case of a Hilbert space regularization, $R(u) = \frac{1}{2}\|u\|^2$, we have

$$E_{\alpha,\zeta}(w; \xi) = \frac{1}{2\zeta} \|\xi - A^* w\|_X^2 + \frac{\alpha}{2} \|w\|_Y^2$$

and the problem of computing the minimizer is a classical Tikhonov regularization problem. In the quadratic case it is easy to see that the two definitions for approximated source condition (3.7) and (3.8) are closely related. This intuition holds also for more general R as we have shown in [II].

In the paper [Theorem 2.9., II] we prove that one can get general estimates for the error between the approximated solution and the true solution in Bregman distance by assuming that $R(u_\alpha^R - u^*)$ can be estimated by the Bregman distance $D_R^{\xi_\alpha,\xi^*}(u_\alpha^R, u^*)$. This assumption turns out to be rather natural as shown in the examples. We have also calculated the convergence rates of the Bregman distance in the case of $R(u) = \frac{1}{p}\|u\|_X^p$ with $p \in [1, \infty)$. This includes the case of one-homogeneous regularization functionals which are usually excluded in the analysis of error estimates.

In the case of one homogeneous regularization functional $R(u) = \|u\|_X$ and the approximated source condition (3.8) written in the form

$$\inf_{w \in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(\xi^* - A^* w) \leq \beta \right\} = C_1 \beta^{-r_1}$$

and

$$\inf_{w \in Y} \left\{ \|w\|_Y^2 \;\middle|\; S(\eta - A^* w) \leq \beta \right\} = C_2 \beta^{-r_2}$$

when $\beta > 0$ small enough we get the following convergence result.

Theorem 2 *Let X be a Banach space and $R(u) = \|u\|_X$. Suppose that the above assumption is satisfied with some orders $r_1, r_2 \geq 0$. For the choice $\alpha \simeq \delta^\kappa$ where*

$$\kappa = \begin{cases} \frac{(1+r_1)(2+r_2)}{(2+r_1)(1+r_2)} & \text{for } r_1 \leq r_2 \text{ and} \\ 1 & \text{for } r_2 < r_1, \end{cases}$$

we have that

$$D_R^{\xi_\alpha,\xi^*}(u_\alpha^R, u^*) \lesssim \begin{cases} \delta^{\frac{2+r_2}{(2+r_1)(1+r_2)}} & \text{for } r_1 \leq r_2 \text{ and} \\ \delta^{\frac{1}{1+r_1}} & \text{for } r_2 < r_1. \end{cases}$$

Notice that in the case when we do not have a strict source condition, the corresponding parameter r_j must be positive.

3.4 Frequentist framework

Our main motivation behind the interest in regularization with large noise is the stochastic noise modelling and especially the statistics of white noise. In the case of random noise, however, the approximate source condition needs to be reconsidered in a statistical framework.

In [II] we define general frequentist risk in Bregman distance between the estimator $U_\alpha^R = U_\alpha^R(\omega)$ and u^* by

$$E_B(U_\alpha^R, u^*) = \mathbb{E}\big(D_R^{\xi_\alpha,\xi^*}(U_\alpha, u^*)\big). \tag{3.9}$$

In quadratic case $R(u) = \|u\|_X^2$ general frequentist risk (3.9) is simply the mean integrated squared error (MISE)

$$E_B(U_\alpha^R, u^*) = \mathbb{E}\|U_\alpha^R - u^*\|_X^2,$$

which implies that E_B is a natural generalization of the often used frequentist risk estimate. Convergence rates of MISE have been widely studied in the literature, see e.g [17, 18] and references therein.

We also observe that a finite estimate can only be obtained if $\mathbb{E}(e_{\delta,\zeta}(A^*\mathcal{E})) < \infty$ at least for some $\zeta > 0$. Under the typical choices of R the finiteness for any δ and ζ is obtained if

$$\mathbb{E}(e_{1,1}(A^*\mathcal{E})) < \infty.$$

This condition can be interpreted as an abstract smoothing condition for the forward operator A. In Gaussian case for example the above assumption is fulfilled if A is a trace-class operator.

In paper [II] we study and derive the convergence rate of frequentist risk for three popular examples: quadratic Tikhonov regularization, Besov norm regularization and total variation regularization. As for the noise we assume the canonical white Gaussian noise model on the Gelfand triplet (Z, Y, Z^*).

4 Bayesian inverse problems

As mentioned in Section 2.2 the maximum a priori estimate of a discrete Bayesian inverse problem with a fixed realization of noise and unknown corresponds to solving the minimization problem

$$\mathbf{u}_\alpha^B := \arg\min_{\mathbf{u}\in\mathbb{R}^n} \left\{ \|\mathbf{A}\mathbf{u} - \mathbf{m}\|_2^2 + \delta^2 \|\mathbf{C}_\mathbf{U}^{-1/2}\mathbf{u}\|_2^2 \right\} \tag{4.1}$$

where $\mathbf{C}_\mathbf{U}$ is the covariance matrix of $\mathbf{U}$. Since the above is just a Tikhonov regularization problem we can handle the difficulties arising from continuous large noise assumption the same way as in section 3.1. If we assume that $\mathbf{C}_\mathbf{U}$ is a discretization of an order $2r$ smoothing pseudodifferential operator C_U then in continuous case we are interested in an estimate

$$u_\alpha^B = \underset{u \in H^r(N)}{\arg\min} \big\{ \|Au\|^2_{L^2(N)} - 2\langle m_\delta, Au\rangle + \delta^2 \|C_U^{-1/2} u\|^2_{L^2(N)} \big\}, \tag{4.2}$$

where $\langle m_\delta, Au\rangle$ is interpreted as a suitable duality pairing instead of $L^2(N)$ inner product. When A is a pseudodifferential operator of order $-t$, where $t \geq s - r$, we can define $\langle m_\delta, Au\rangle = \langle m_\delta, Au\rangle_{H^{-s}(N)\times H^s(N)}$. Note that the forward operator A, the prior distribution and the noise depend on on each other only through assumption $t \geq s - r$.

If we are thinking the above as a MAP estimate to a Bayesian problem we have to assume that U has formally the following distribution

$$\pi_{pr}(u) \underset{formally}{=} c \exp\Big(-\frac{1}{2} \|C_U^{-1/2} u\|^2_{L^2(N)} \Big).$$

To see why the above distribution is defined only formally let us take a simple example in 1-dimensional torus $\mathbb{T}^1$. We are interested in the inverse problem

$$M_\delta = AU + \delta \mathcal{E}$$

where we assume that $\mathcal{E} \sim N(0, I)$ and $U \sim N(0, I)$, that is both the noise and the unknown are assumed to be normalised white Gaussian noise. As we have seen before white noise takes values in $H^\tau(\mathbb{T}^1)$ with some $\tau < -1/2$. On the other hand white noise has formally the following distribution

$$\pi_{pr}(u) \underset{formally}{=} c \exp\Big(-\frac{1}{2} \|u\|^2_{L^2(\mathbb{T}^1)} \Big).$$

Hence we want to solve

$$\underset{u \in L^2}{\arg\min} \big\{ \|Au\|^2_{L^2(\mathbb{T}^1)} - 2\langle m_\delta, Au\rangle + \delta^2 \|u\|^2_{L^2(\mathbb{T}^1)} \big\}.$$

Note carefully that we are looking for a solution in $L^2(\mathbb{T}^1)$ even though the realizations of U are in $L^2(\mathbb{T}^1)$ with probability zero. In general if we are interested in finding a solution in $H^r(N)$, which means that we need to choose a covariance operator C_U that is $2r$ orders smoothing, then we can show that the prior takes values in $H^\tau(N)$ where $\tau = r - s$, see [III].

The main interest of the paper [III] lies in the convergence of

$$\mathbb{E}\|U_\delta^B(\omega) - U(\omega)\|_{H^\zeta} \qquad \text{when } \delta \to 0. \tag{4.3}$$

Here $\zeta < \tau = r - s$ and the expectation is taken with respect to the joint distribution of $(U, \mathcal{E})$. We can prove a similar theorem about the convergence rates of (4.3) as in the deterministic case in Section 3.1. The convergence rates and the upper limit of ζ turn out to be same as in the deterministic approach whit choice $\kappa = 2$.

We notice that in the Bayesian case the upper limit $\zeta < \tau$ for the smoothness index of the space of convergence makes trivially sense since when $\zeta > \tau$ a random draw U from the prior distribution is in $H^\zeta(N)$ with probability zero. This means that even when we study Tikhonov regularization from a purely deterministic point of view the regularized solution still inherits some behaviour that is easiest explained by interpreting it as a MAP estimate of Gaussian Bayesian inverse problem. On the other hand deterministic regularization gives us the liberty of choosing α freely. Since in the purely Bayesian approach the prior information should be independent of the measurement M_δ the regularization parameter α is determined by the variance of the noise $\delta\mathcal{E}$ and hence $\alpha = \delta^2$. In literature this principle is occasionally omitted and general a priori rules $\alpha = \alpha(\delta)$ are considered. Such an approach resembling the frequentist method leads to 'priors' that are scaled with respect to the noise level δ and hence no longer independent of the measurement. With general $\alpha(\delta)$ the minimization problem (4.2) can not be seen as a proper MAP estimate. However, in many cases it is a useful estimator to study.

4.1 Uncertainty quantification

One advantage Bayesian inversion offers over deterministic regularization is uncertainty quantification. Since the solution to the Bayesian inverse problem is the posterior distribution of the unknown we can study its credible sets and their contraction in some Sobolev space H^ζ when $\delta \to 0$. A Bayesian credible set is a region in the posterior distribution that contains a large fraction of the posterior mass, for instance 95%. We are dealing with Gaussian distributions which are symmetric, so we define our credible sets to be central regions. This means these sets are defined as central balls with centre u_δ^B.

These credible sets are often used to visualize the remaining *Bayesian uncertainty* in the estimate. Frequentists use another kind of uncertainty quantification called confidence region. A confidence region is a range of values that frequently includes the unknown of interest if the experiment is repeated. We can define confidence regions as central balls with $u_\delta^\dagger$ as the centre. Here $u_\delta^\dagger$ is the frequentist approximated solution generated by a true solution $u^\dagger$. How frequently the ball around the approximated solution, with different realization of the noise, contains the true solution is determined by the confidence level. Note that whether one confidence region covers the unknown $u^\dagger$ or not is no longer a matter of probability. This means that a 95% confidence region does not imply that, for a given realized region calculated from the measurement data, there is a 95% probability the true solution lies within the region. Nor does it mean that there is a 95% probability that the region covers the true solution.

In the finite-dimensional parametric case and under mild conditions on the prior, the Bernstein–von Mises Theorem implies that the credible sets of smooth models are asymptotically equivalent with the frequentist confidence regions based on the maximum likelihood estimator, see [75]. In the infinite-dimensional case there is no corresponding theorem and hence Bayesian credible sets are not automatically frequentist confidence sets. This means that if we assume that the data is generated by a true parameter $u^\dagger$, it is not automatically true that credible sets contain that truth with probability at least the credible level. However the correspondence of Bayesian and frequentist uncertainty has been studied in many recent papers, for instance in [15, 16, 47, 51, 60, 71]. These results are important since they show that some credible sets nicely illustrate the uncertainty of the estimate in the classical frequentist sense.

In paper [III] we show that the posterior covariance operator converges which, with the convergence of the mean, guarantees the weak convergence of the posterior distribution. We also study both Bayesian credible sets and frequentist confidence regions. If we assume that $\tau > 0$ we can prove the following posterior contraction in the frequentist setting

$$\mathbb{E}_{u^\dagger}\mathbb{P}_{M_\delta^\dagger}\Big\{u \in H^\tau(N)\,\big|\, \|u-u^\dagger\|_{L^2(N)} \geq c\delta^\kappa\Big\} \to 0$$

when $\delta \to 0$ for all $c > 0$ and $\kappa < \frac{2\tau}{s+\tau+t}$. Since $s = \frac{d}{2} + \epsilon$ the above convergence rate agrees, up to $\epsilon > 0$ arbitrarily small, with the minimax convergence rate, see [17].

5 Conclusions

This thesis presents convergence and contraction results for infinite dimensional linear inverse problems when large noise is assumed. The problem is studied in regularization, frequentist and Bayesian settings and hence covers the usual methods used to solve inverse problems. Our interest lies in indirect measurement corrupted by large noise. The discrete version of such models is used in countless practical inverse problems. Connecting discrete a model to an infinite-dimensional limit model is desirable since such a connection provides, for instance, discretization invariant error analysis for numerical inversion and computational efficiency based on robust switching between different discretizations related to multigrid methods.

In paper [I] we study Tikhonov regularization when the noise is assumed to be a realization of white Gaussian noise. The focus of our analysis is the apparent paradox arising from the infinite L^2-norm of the natural limit of white Gaussian noise in $\mathbb{R}^k$ when $k \to \infty$. We show how to build a rigorous theory removing this paradox and we explain how to take this into account in discrete inverse problems using appropriate Sobolev scales. We prove that with the correct choice of regularization parameter the approximated solution converges in a space with smoothness index arbitrarily close to the smoothness of the true solution. We also note that since much larger noise is allowed than in the classical theory we need stronger regularization to achieve convergence.

In paper [II] we take a more general approach and tackle the issue of large noise regularization with convex regularization functionals in Banach spaces. We derive a rather general theory that can be adapted to special homogeneity properties of the regularization functional, in particular to quadratic and one-homogeneous regularizations. The first one corresponds to Tikhonov regularization whereas the second one was popularized via total variation methods. Our key contribution is to derive Bregman distance-based error estimates for the regularized solution with convex regularization functional. Given a deterministic model for large noise, one can derive explicit converge rate results given an approximate source condition for the unknown and noise. We also prove convergence results for general frequentist risk where the error measure is given by the Bregman distance.

The goal of paper [III] is to use Bayesian inversion to construct a consistent continuous-discrete framework covering the case of white Gaussian noise in statistical inverse problems. Developing such a theory is important since analytic results about the continuous model can then be restricted to a given resolution in a discretization invariant way. We are then interested to know what happens to the posterior distribution when the noise amplitude goes to zero. This analysis of small noise limit, also known as the theory of posterior consistency, has attracted a lot of interest during the past decade. In many earlier studies the forward operator and the covariance operators of the noise and the prior are assumed to be simultaneously diagonalizable. In more recent studies this is not required but the operators are tied to each other in a highly non-trivial way. The notable advance of the approach in [III] compared to previous studies is that the forward operator, the prior distribution and the noise depend on each other only through the assumption that the forward operator is smoothing enough.

References

[1] Sergios Agapiou, Stig Larsson, and Andrew M Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. *Stochastic Processes and their Applications*, 123(10):3828–3860, 2013.

[2] Sergios Agapiou, Andrew M Stuart, and Yuan-Xiang Zhang. Bayesian posterior contraction rates for linear severely ill-posed inverse problems. *Journal of Inverse and Ill-posed Problems*, 22(3):297–321, 2014.

[3] Martin Benning and Martin Burger. Error estimates for general fidelities. *Electronic Transactions on Numerical Analysis*, 38(44-68):77, 2011.

[4] N. Bissantz, T. Hohage, and A. Munk. Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise. *Inverse Problems*, 20:1773–1789, 2004.

[5] George EP Box, J Stuart Hunter, and William Gordon Hunter. *Statistics for experimenters: design, innovation, and discovery*, volume 2. Wiley-Interscience New York, 2005.

[6] Andrea Braides. *Gamma-convergence for Beginners*, volume 22. Clarendon Press, 2002.

[7] Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

[8] Alfred M Bruckstein, David L Donoho, and Michael Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM review*, 51(1):34–81, 2009.

[9] Martin Burger. Bregman distances in inverse problems and partial differential equation. *arXiv preprint arXiv:1505.05191*, 2015.

[10] Martin Burger and Andreas Neubauer. Analysis of Tikhonov regularization for function approximation by neural networks. *Neural Networks*, 16(1):79–90, 2003.

[11] Martin Burger and Stanley Osher. A guide to the TV zoo. In *Level Set and PDE Based Reconstruction Methods in Imaging*, pages 1–70. Springer, 2013.

[12] Martin Burger, Elena Resmerita, and Lin He. Error estimation for Bregman iterations and inverse scale space methods in image restoration. *Computing*, 81(2-3):109–135, 2007.

[13] D. Calvetti and E. Somersalo. *Introduction to Bayesian scientific computing: ten lectures on subjective computing*, volume 2. Springer, 2007.

[14] George Casella and Roger L Berger. *Statistical inference*, volume 2. Duxbury Pacific Grove, CA, 2002.

[15] Ismaël Castillo and Richard Nickl. Nonparametric Bernstein–von Mises theorems in Gaussian white noise. *The Annals of Statistics*, 41(4):1999–2028, 2013.

[16] Ismaël Castillo and Richard Nickl. On the Bernstein–von Mises phenomenon for nonparametric Bayes procedures. *The Annals of Statistics*, 42(5):1941–1969, 2014.

[17] L Cavalier. Nonparametric statistical inverse problems. *Inverse Problems*, 24(3):034004, 2008.

[18] Laurent Cavalier and Alexandre Tsybakov. Sharp adaptation for inverse problems with random noise. *Probability Theory and Related Fields*, 123(3):323–354, 2002.

[19] Simon L Cotter, Massoumeh Dashti, James Cooper Robinson, and Andrew M Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. *Inverse problems*, 25(11):115008, 2009.

[20] Masoumeh Dashti, Kody JH Law, Andrew M Stuart, and Jochen Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. *Inverse Problems*, 29(9):095017, 2013.

[21] Ingrid Daubechies, Gerd Teschke, and Luminita Vese. Iteratively solving linear inverse problems under general convex constraints. *Inverse Problems and Imaging*, 1(1):29, 2007.

[22] Paul N Eggermont, Vincent LaRiccia, and M Zuhair Nashed. Noise models for ill-posed problems. In *Handbook of Geomathematics*, pages 739–762. Springer, 2010.

[23] PPB Eggermont, VN LaRiccia, and MZ Nashed. On weakly bounded noise in ill-posed problems. *Inverse Problems*, 25(11):115018, 2009.

[24] PPB Eggermont, VN LaRiccia, and MZ Nashed. Moment discretization for ill-posed problems with discrete weakly bounded noise. *GEM-International Journal on Geomathematics*, 3(2):155–178, 2012.

[25] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[26] S.N. Evans and P.B. Stark. Inverse problems as statistics. *Inverse Problems*, 18:R55–R97, 2002.

[27] B.G. Fitzpatrick. Bayesian analysis in inverse problems. *Inverse problems*, 7:675–702, 1991.

[28] J.N. Franklin. Well posed stochastic extensions of ill posed linear problems. *Journal of the Institute of Mathematics and Its Applications*, 31:682–716, 1970.

[29] Manuel Freiberger, Christian Clason, and Hermann Scharfetter. Total variation regularization for nonlinear fluorescence tomography with an augmented Lagrangian splitting approach. *Applied optics*, 49(19):3741–3747, 2010.

[30] Subhashis Ghosal, Jayanta K Ghosh, and Aad W Van Der Vaart. Convergence rates of posterior distributions. *Annals of Statistics*, 28(2):500–531, 2000.

[31] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of l^1-regularization. *Communications on Pure and Applied Mathematics*, 64(2):161–182, February 2011.

[32] Allan Greenleaf and Gunther Uhlmann. Microlocal techniques in integral geometry. *Integral geometry and tomography (Arcata, CA, 1989)*, 113:121–135, 1990.

[33] C. W. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind*, volume 105 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[34] J. Hadamard. Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton Univ. Bull.*, 13, 1902.

[35] Torsten Hein. Convergence rates for regularization of ill-posed problems in Banach spaces by approximate source conditions. *Inverse Problems*, 24(4):045007, 2008.

[36] Tapio Helin. Discretization and Bayesian modeling in inverse problems and imaging. *Academic dissertation, Aalto University*, 2010.

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

[38] Bernd Hofmann. Approximate source conditions in Tikhonov–Phillips regularization and consequences for inverse problems with multiplication operators. *Mathematical Methods in the Applied Sciences*, 29(3):351–371, 2006.

[39] Bernd Hofmann and Masahiro Yamamoto. Convergence rates for Tikhonov regularization based on range inclusions. *Inverse Problems*, 21(3):805, 2005.

[40] Lars Hörmander. *The analysis of linear partial differential operators. III*, volume 274 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1994. Pseudo-differential operators, Corrected reprint of the 1985 original.

[41] Tzee-Ming Huang. Convergence rates for posterior distributions and adaptive estimation. *The Annals of Statistics*, 32(4):1556–1593, 2004.

[42] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.

[43] Stefan Kindermann. Convex Tikhonov regularization in Banach spaces: New results on convergence rates. *Journal of Inverse and Ill-posed Problems*, 2016.

[44] Andreas Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[45] B. T. Knapik, B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Related Fields*, 164(3-4):771–813, 2016.

[46] Bartek Knapik and Jean-Bernard Salomond. A general approach to posterior contraction in nonparametric inverse problems. *arXiv preprint arXiv:1407.0335*, 2014.

[47] BT Knapik, AW van Der Vaart, and JH Van Zanten. Bayesian inverse problems with Gaussian priors. *The Annals of Statistics*, 39(5):2626–2657, 2011.

[48] BT Knapik, AW Van der Vaart, and JH Van Zanten. Bayesian recovery of the initial condition for the heat equation. *Communications in Statistics-Theory and Methods*, 42(7):1294–1313, 2013.

[49] V. Kolehmainen, S. Siltanen, S. Järvenpää, JP Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical X-ray tomography with few radiographs: II. application to dental radiology. *Physics in Medicine and Biology*, 48:1465–1490, 2003.

[50] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging*, 3:87–122, 2009.

[51] Haralambie Leahu. On the Bernstein-von Mises phenomenon in the Gaussian white noise model. *Electronic Journal of Statistics*, 5:373–404, 2011.

[52] Kui Lin, Shuai Lu, and Peter Mathé. Oracle-type posterior contraction rates in Bayesian inverse problems. *Inverse Problems & Imaging*, 9(3), 2015.

[53] D. A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-posed Problems*, 16(5):463–478, 2008.

[54] Shuai Lu and Sergei V Pereverzev. *Regularization theory for ill-posed problems: selected topics*, volume 58. Walter de Gruyter, 2013.

[55] Dennis McLaughlin and Lloyd R Townley. A reassessment of the groundwater inverse problem. *Water Resources Research*, 32(5):1131–1161, 1996.

[56] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer-Verlag, New York, 1984. Translated from the Russian by A. B. Aries, Translation edited by Z. Nashed.

[57] David L Phillips. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM (JACM)*, 9(1):84–97, 1962.

[58] Ronny Ramlau and Gerd Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numerische Mathematik*, 104:177–203, 2006. 10.1007/s00211-006-0016-3.

[59] Kolyan Ray. Bayesian inverse problems with non-conjugate priors. *Electronic Journal of Statistics*, 7:2516–2549, 2013.

[60] Kolyan Ray. Adaptive bernstein-von mises theorems in gaussian white noise. *arXiv:1407.3397*, 2014.

[61] E. Resmerita. Regularization of ill-posed problems in Banach spaces: convergence rates. *Inverse Problems*, 21:1303, 2005.

[62] Elena Resmerita and Otmar Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801, 2006.

[63] Ju. A. Rozanov. *Infinite-dimensional Gaussian distributions*. American Mathematical Society, Providence, R. I., 1971. Translated from the Russian by G. Biriuk.

[64] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[65] Thomas Schuster, Barbara Kaltenbacher, Bernd Hofmann, and Kamil S. Kazimierski. *Regularization methods in Banach spaces*, volume 10 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2012.

[66] Xiaotong Shen and Larry Wasserman. Rates of convergence of posterior distributions. *Ann. Statist.*, 29(3):687–714, 2001.

[67] M. A. Shubin. *Pseudodifferential operators and spectral theory*. Springer-Verlag, Berlin, second edition, 2001. Translated from the 1978 Russian original by Stig I. Andersson.

[68] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical X-ray tomography with few radiographs: I. general theory. *Physics in medicine and biology*, 48:1437–1463, 2003.

[69] Andrew M Stuart. Inverse problems: a Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[70] VN Sudakov and LA Khalfin. Statistical approach to correctness of problems in mathematical physics. *DOKLADY AKADEMII NAUK SSSR*, 157(5):1058, 1964.

[71] Botond Szabo, Aad van der Vaart, and Harry van Zanten. Honest Bayesian confidence sets for the L2-norm. *Journal of Statistical Planning and Inference*, 166:36–51, 2015.

[72] Botond Szabó, AW van der Vaart, and JH van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428, 2015.

[73] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial Mathematics, 2005.

[74] A. N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, and A. G. Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1995. Translated from the 1990 Russian original by R. A. M. Hoksbergen and revised by the authors.

[75] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.

[76] Curtis R. Vogel. *Computational methods for inverse problems*, volume 23 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002. With a foreword by H. T. Banks.

[77] Sebastian J Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Problems*, 29(12):125011, 2013.