

Chapter 5

Inverse Problems



Thorsten Hohage, Benjamin Sprung and Frederic Weidling

The shortest path between two truths in the real domain passes through the complex domain.
—Jacques Hadamard

5.1 Introduction

5.1.1 What Is an Inverse Problem?

Generally speaking, inverse problems typically consist in the reconstruction of causes for observed effects. In imaging applications the cause is usually a probe and the effect are observed data. The corresponding forward problems then consists in predicting experimental data given perfect knowledge of the probe. In some sense solving an inverse problems means “computing backwards”, which is usually more difficult then solving the forward problem.

To model these kind of problems mathematically we describe the imaging system or experimental setup by a forward operator $F: \mathbb{X} \rightarrow \mathbb{Y}$ between Banach spaces \mathbb{X}, \mathbb{Y} , which maps a probe $f \in \mathbb{X}$ to the corresponding effect $g \in \mathbb{Y}$. Then the inverse problem is, given data $g \in \mathbb{Y}$, to find a solution $f \in \mathbb{X}$ to the equation

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

T. Hohage (✉) · B. Sprung · F. Weidling
Institute for Numerical and Applied Mathematics, Universität Göttingen, Lotzestr. 16-18, 37083
Göttingen, Germany
e-mail: hohage@math.uni-goettingen.de

B. Sprung
e-mail: b.sprung@math.uni-goettingen.de

F. Weidling
e-mail: f.weidling@math.uni-goettingen.de

5.1.2 Ill-Posedness and Regularization

A first obvious question to ask is whether or not a probe f is uniquely determined by the data g , i.e. if the operator F is injective. Such questions are addressed e.g. in Sect. 13.6.6. Given uniqueness, one might try to find the solution f to (5.1) by just applying the inverse operator F^{-1} , which gives another reason to call the problem an inverse problem. However in practice this can cause several problems, due to the forward operator being *ill-posed*.

According to J. Hadamard a problem is called *well-posed* if the following conditions are satisfied:

1. There exists a solution.
2. The solution is unique.
3. The solution depends continuously on the data (stability).

Otherwise the problem is called *ill-posed*.

An inverse problem in the form of an operator equation (5.1) is well-posed if F is surjective (such that for all g there exists a solution), injective (such that the solution is unique) and if F^{-1} is continuous (guaranteeing stability). For many inverse problems in practice only the third condition is violated, and ill-posedness in the narrower sense often refers to this situation: The reconstruction of causes from observed effects is unstable since very different causes may have similar effects.

The remedy against ill-posedness is *regularization*: To obtain stable solutions of inverse problems one constructs a family of continuous operators $R_\alpha : \mathbb{Y} \rightarrow \mathbb{X}$ parameterized by a parameter $\alpha > 0$ converging pointwise to the discontinuous inverse F^{-1} :

$$R_\alpha \approx F^{-1}, \quad \lim_{\alpha \rightarrow 0} R_\alpha(F(f)) = f \quad \text{for all } f \in \mathbb{X}. \quad (5.2)$$

We will discuss several generic constructions of such families of stable approximate inverses R_α in §5.2.

5.1.3 Examples

5.1.3.1 Numerical Differentiation

In our first example we consider the forward operator given by integration. We fix the free integration constant by working in spaces of functions with mean zero, $L^2_\diamond([0, 1]) := \{f \in L^2([0, 1]) : \int_0^1 f(x) dx = 0\}$. Let $F : L^2_\diamond([0, 1]) \rightarrow L^2_\diamond([0, 1])$ be given by

$$F(f)(y) = \int_0^y f(x) dx + c(f), \quad y \in [0, 1]. \quad (5.3)$$

where $c(f)$ is such that $F(f) \in L^2_\diamond$. The corresponding inverse problem described by the operator equation (5.1) is to compute the derivative g' . If $g = F(f)$, then the existence of a solution to the inverse problem is guaranteed and this solution is unique. Now assume that instead of the exact data g we are given noisy data g^{obs} that fulfill

$$\|g^{\text{obs}} - g\|_{L^2} \leq \delta,$$

for some small noise level $\delta > 0$. For example this could be one of the functions

$$g_n^{\text{obs}}(x) = g(x) + \delta\sqrt{2} \sin(\pi nx).$$

As the derivatives are given by $(g_n^{\text{obs}})'(x) = g'(x) + \pi n\delta\sqrt{2} \cos(\pi nx)$, we have

$$\|(g_n^{\text{obs}})' - g'\|_{L^2} = \pi n\delta.$$

This illustrates the typical ill-posedness of inverse problems: amplification of noise may be arbitrarily large for naive application of the inverse F^{-1} . This is the main difficulty one has to cope with. Our toy example illustrates another typical feature of inverse problems: Each function in the image of the operator F defined above is at least once differentiable. Also for many other inverse problems the forward operator is smoothing in the sense that the output function has higher smoothness than the input function, and this property causes the instability of the inverse problem.

5.1.3.2 Fluorescence Microscopy

In fluorescence microscopy one is interested in recovering the density f of fluorescent markers in some specimen in \mathbb{R}^3 . The probe is sampled by a laser beam, and one detects fluorescent photons. In confocal microscopy, spatial resolution is achieved by focusing the laser beam by some lens and collecting fluorescent electrons by same lens such that out-of-focus fluorescent photons can be blocked by a pinhole.

Let $y \in \mathbb{R}^3$ be the focal point and assume that the probability (density) that for the focus point $y \in \mathbb{R}^3$ a fluorescent photon emitted by a marker at point $x \in \mathbb{R}^3$ is detected is $k(x - y)$. k is called the *point-spread function*, and we assume here that it is spatially invariant. Then our problem is described by the operator equation

$$g(y) = F(f)(y) = \int k(y - x)f(x) dx,$$

i.e. the observation g is given by a convolution of the marker density f with point spread function k . As convolution will usually blur an image, the forward operator is smoothing. Smoother kernels will lead to stronger smoothing.

5.1.4 Choice of Regularization Parameters and Convergence Concepts

Due to the ill-posedness discussed above it is essential to take into account the effects of noise in the observed data. Let $f^\dagger \in \mathbb{X}$ denote the unknown exact solution, and first assume data g^δ with deterministic errors such that

$$\|g^\delta - F(f^\dagger)\|_{\mathbb{Y}} \leq \delta. \quad (5.4)$$

As mentioned above, regularization of an ill-posed operator equation (5.1) with an injective operator F consists in approximating the discontinuous inverse operator F^{-1} by a pointwise convergent family of continuous operators $R_\alpha : \mathbb{Y} \rightarrow \mathbb{X}, \alpha > 0$. This immediately gives rise to the question which operator in the family should be chosen for the reconstruction, i.e how to choose the parameter α . Usually the starting point of deterministic error analysis in regularization theory is the following splitting of the reconstruction error:

$$\|R_\alpha(g^\delta) - f^\dagger\| \leq \|R_\alpha(g^\delta) - R_\alpha(F(f^\dagger))\| + \|R_\alpha(F(f^\dagger)) - f^\dagger\|. \quad (5.5)$$

The first term on the right hand side is called *propagated data noise error*, and the second term is referred to as *approximation error* or *bias*. Due to pointwise convergence (see (5.2)), the bias tends to 0 as $\alpha \rightarrow 0$. Hence, to control this error term, we should choose α as small as possible. However, as R_α converges pointwise to the discontinuous operator F^{-1} , the Lipschitz constant (or operator norm in the linear case) of R_α will explode as $\alpha \rightarrow 0$, and hence also the propagated data noise error. Therefore, α must not be chosen too small. This indicates that the choice of the regularization parameter must be a crucial ingredient of a reliable regularization method. Probably the most well-known parameter choice rule in the deterministic setting is *Morozov's discrepancy principle*:

$$\bar{\alpha}_{\text{DP}}(\delta, g^\delta) := \sup\{\alpha > 0 : \|F(R_\alpha(g^\delta)) - g^\delta\| \leq \tau\delta\} \quad (5.6)$$

with some parameter $\tau \geq 1$. In other words, among all estimators $R_\alpha(g^\delta)$ which can explain the data within the noise level (times τ), we choose the most stable one.

Definition 5.1 A family of operators $R_\alpha : \mathbb{Y} \rightarrow \mathbb{X}$ parameterized by a parameter $\alpha > 0$ together with some rule $\bar{\alpha} : [0, \infty) \times \mathbb{Y} \rightarrow (0, \infty)$ how to choose this parameter depending on the noise level δ and the data g^δ is called a *regularization method* if the worst case error tends to 0 with the noise level in the sense that

$$\lim_{\delta \rightarrow 0} \sup \{ \|R_{\bar{\alpha}(\delta, g^\delta)}(g^\delta) - f^\dagger\|_{\mathbb{X}} : g^\delta \in \mathbb{Y}, \|g^\delta - F(f^\dagger)\|_{\mathbb{Y}} \leq \delta \} = 0 \quad (5.7)$$

for all $f^\dagger \in \mathbb{X}$.

The convergence (5.7) is a minimal requirement that one expects from a regularization method. However, it can be shown that for ill-posed problems this convergence may be arbitrarily slow depending on f^\dagger . This is of course not satisfactory. Fortunately, a-priori information on the solution f^\dagger , which is often available, may help. If we know a-priori that f^\dagger belongs to some set $\mathcal{K} \subset \mathbb{X}$, then it is often possible to derive explicit error bounds

$$\sup \left\{ \|R_{\bar{\alpha}(\delta, g^\delta)}(g^\delta) - f^\dagger\|_{\mathbb{X}} : g^\delta \in \mathbb{Y}, \|g^\delta - F(f^\dagger)\|_{\mathbb{Y}} \leq \delta \right\} \leq \psi(\delta)$$

for all $f^\dagger \in \mathcal{K}$ with a function $\psi \in C([0, \infty))$ satisfying $\psi(0) = 0$.

Let us now consider statistical noise models instead of the deterministic noise model (5.4). Often statistical data G_t belong to a different space \mathbb{Y}' , e.g. a space of distributions. The distribution depends on some parameter t , and we assume that $G_t \rightarrow F(f^\dagger)$ in some sense as $t \rightarrow \infty$. Let t e.g. denote the number of observations or in photonic imaging the expected number of photons. As the estimator $R_\alpha(g^{\text{obs}})$ (where now R_α is a mapping from \mathbb{Y}' to \mathbb{X}) will be a random variable, we have to use stochastic concepts of convergence, e.g. convergence in expectation. Other convergence concepts, in particular convergence in probability are also used frequently.

Definition 5.2 In the setting above, a family of operators $R_\alpha : \mathbb{Y}' \rightarrow \mathbb{X}$ parameterized by a parameter $\alpha > 0$ together with some parameter choice rule $\bar{\alpha} : [0, \infty) \times \mathbb{Y}' \rightarrow (0, \infty)$ is called a *consistent estimator* if

$$\lim_{t \rightarrow \infty} \mathbb{E} \left[\|R_{\bar{\alpha}(t, G_t)}(G_t) - f^\dagger\|_{\mathbb{X}}^2 \right] = 0$$

for all $f^\dagger \in \mathbb{X}$.

Again one may further ask not only for convergence, but even rates of convergence as $t \rightarrow \infty$ on certain subsets $\mathcal{K} \subset \mathbb{X}$.

5.2 Regularization Methods

In this section we will discuss generalized Tikhonov regularization, which is given by finding the minimum of

$$\hat{f}_\alpha \in \operatorname{argmin}_{f \in \mathbb{X}} \left[\mathcal{S}_{g^{\text{obs}}}(F(f)) + \alpha \mathcal{R}(f) \right], \quad (5.8)$$

where $\mathcal{S}_{g^{\text{obs}}}$ is the *data fidelity functional*, which measures some kind of distance of $F(f)$ and the data g^{obs} and causes the minimizer \hat{f}_α to still explain the data well, whereas \mathcal{R} is the *penalty functional* which penalizes certain properties of the minimizer. This approach is called *variational regularization*, as our regularized solution is found by minimization of a functional (usually an integral functional).

5.2.1 Variational Regularization

We start with a probabilistic motivation of generalized Tikhonov regularization. From here until the end of this section we will consider the finite-dimensional setting $\mathbb{X} = \mathbb{R}^n$, $\mathbb{Y} = \mathbb{R}^m$ and use boldface symbols to denote finite-dimensional vectors/mappings. We start from (5.1), where $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{g} \in \mathbb{R}^m$ and $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is some injective function. Given the data \mathbf{g} we want to find the solution \mathbf{f} of $F(\mathbf{f}) = \mathbf{g}$, but recall that we cannot just apply the inverse F^{-1} as discussed in Sect. 5.1. Instead we might estimate \mathbf{f} by maximizing the likelihood function $\mathcal{L}(\mathbf{f}) = P(\mathbf{g}|\mathbf{f})$, i.e. the probability that for a certain preimage \mathbf{f} the data \mathbf{g} will occur. If we assume that our data is normally distributed with covariance matrix $\sigma^2 I$, then we can rearrange the problem by using the monotonicity of the logarithm, as well as the fact, that neither additive nor multiplicative constants change the extremal point:

$$\begin{aligned} \hat{\mathbf{f}}_{\text{ML}} &\in \operatorname{argmax}_{\mathbf{f} \in \mathbb{R}^n} P(\mathbf{g}|\mathbf{f}) = \operatorname{argmax}_{\mathbf{f} \in \mathbb{R}^n} \log(P(\mathbf{g}|\mathbf{f})) \\ &= \operatorname{argmax}_{\mathbf{f} \in \mathbb{R}^n} \log \left(\frac{1}{\sqrt{2\pi}\sigma} \prod_{i=1}^m \exp \left(\frac{-(\mathbf{g}_i - F(\mathbf{f})_i)^2}{2\sigma^2} \right) \right) \\ &= \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \frac{1}{2\sigma^2} \sum_{i=1}^m (\mathbf{g}_i - F(\mathbf{f})_i)^2 = \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \frac{1}{2\sigma^2} \|\mathbf{g} - F(\mathbf{f})\|_2^2. \end{aligned}$$

This demonstrates the well-known fact that the maximum likelihood approach for Gaussian noise yields the least squares methods as first used by young Gauss to predict the path of the asteroid Ceres in 1801. However, as $\hat{\mathbf{f}}_{\text{ML}} = F^{-1}(\mathbf{g})$ for \mathbf{g} in the range of F , this approach has no regularizing effect. In fact it is more reasonable to maximize $P(\mathbf{f}|\mathbf{g})$ instead of $P(\mathbf{g}|\mathbf{f})$, as our goal should be to find the solution \mathbf{f} which is most likely to have caused the observation \mathbf{g} , instead of just finding any \mathbf{f} which causes the observation \mathbf{g} with maximal probability. This leads to the Bayesian perspective on inverse problem with the characteristic feature that prior to the measurements a probability distribution (the so-called *prior distribution*) is assigned to the solution space \mathbb{X} modeling our prior knowledge on \mathbf{f} . By Bayes' theorem we have

$$P(\mathbf{f}|\mathbf{g}) = \frac{P(\mathbf{g}|\mathbf{f})P(\mathbf{f})}{P(\mathbf{g})} \Leftrightarrow \text{posterior} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}.$$

Estimating \mathbf{f} by maximizing the posterior $P(\mathbf{f}|\mathbf{g})$ is called maximum a posteriori probability (MAP) estimate. To use this approach we have to model the prior $P(\mathbf{f})$. If we assume that \mathbf{f} is normally distributed with mean $\mathbf{f} \in \mathbb{R}^n$ and covariance matrix $\tau^2 I$, then we find

$$\begin{aligned}
\hat{\mathbf{f}}_{\text{MAP}} &\in \operatorname{argmax}_{\mathbf{f} \in \mathbb{R}^n} P(\mathbf{f}|\mathbf{g}) = \operatorname{argmax}_{\mathbf{f} \in \mathbb{R}^n} [\log(P(\mathbf{g}|\mathbf{f})) + \log(P(\mathbf{f}))] \\
&= \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \frac{1}{2\sigma^2} \left[\sum_{i=1}^m (\mathbf{g}_i - F(\mathbf{f})_i)^2 + \frac{\sigma^2}{\tau^2} \sum_{j=1}^m (\mathbf{f}_j - (\mathbf{f}_0)_j)^2 \right] \\
&= \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \left[\underbrace{\frac{1}{2} \|\mathbf{g} - F(\mathbf{f})\|_2^2 + \frac{\alpha}{2} \|\mathbf{f} - \mathbf{f}_0\|_2^2}_{J_\alpha(\mathbf{f})} \right]
\end{aligned}$$

where $\alpha = \frac{\sigma^2}{\tau^2}$. The functional $J_\alpha(\mathbf{f})$ is the standard (quadratic) Tikhonov functional, and therefore MAP and Tikhonov regularization coincide in this setting.

In photonic imaging the data is often given by photon counts, and these are typically Poisson distributed in the absence of read-out error (compare Chap.4). Recall that a random variable $Z \in \mathbb{N}_0$ is called Poisson distributed with mean $\lambda > 0$, short $Z \sim \text{Pois}(\lambda)$, if $P(Z = g) = e^{-\lambda} \lambda^g / (g!)$ for all $g \in \mathbb{N}_0$. Hence the negative log-likelihood function is given by

$$-\log P(Z = g|\lambda) = \lambda - g \log(\lambda) + \log g! = \lambda - g + g \log\left(\frac{g}{\lambda}\right) + C_g$$

where C_g is a constant independent of λ . Now assume that \mathbf{g} is a vector of independent Poisson distributed random variables such that $\mathbf{g}_i \sim \text{Pois}(F(\mathbf{f})_i)$. It follows that the negative log-likelihood $-\log(P(\mathbf{g}|\mathbf{f})) = -\sum_i \log P(\mathbf{g}_i|F(\mathbf{f})_i)$ is given (up to an additive constant independent of \mathbf{f}) by the Kullback-Leibler divergence

$$\text{KL}(\mathbf{g}, F(\mathbf{f})) := \sum_{i=1}^m \left[F(\mathbf{f})_i - \mathbf{g}_i + \mathbf{g}_i \log\left(\frac{\mathbf{g}_i}{F(\mathbf{f})_i}\right) \right].$$

If \mathbf{f} has a probability distribution $P(f) = c \exp(-\mathcal{R}(\mathbf{f})/\tau^2)$ this leads to generalized Tikhonov regularization

$$\hat{\mathbf{f}}_\alpha \in \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} [\mathcal{S}_\mathbf{g}(F(\mathbf{f})) + \alpha \mathcal{R}(\mathbf{f})] \quad (5.9)$$

with fidelity term $\mathcal{S}_\mathbf{g} = \|\mathbf{g} - \cdot\|_2^2$ for normally distributed data, $\mathcal{S}_\mathbf{g} = \text{KL}(\mathbf{g}, \cdot)$ for Poisson distributed data and the penalty term $\alpha \mathcal{R}$ with regularization parameter $\alpha = \tau^{-2}$ for Poisson data and $\alpha = \sigma^2/\tau^2$ for Gaussian white noise.

Note that in the Bayesian setting above the regularization parameter α is uniquely determined by the prior distribution and the likelihood functional. However, often only qualitative prior knowledge on the solution is available, but the parameter τ is unknown. Then α has to be determined by a-posteriori parameter choice rules analogous to the discrepancy principle (5.6) for deterministic errors.

Let us discuss a few popular choices of the penalty functional \mathcal{R} , which allows the incorporation of prior information on the solution or is simply the negative logarithm of the density of the prior in the above Bayesian setting. For a-priori known sparsity

of the solution one should choose the sparsity enforcing penalty $\mathcal{R}(\mathbf{f}) = \|\mathbf{f}\|_1$. If \mathbf{f} is an image with sharp edges, then the total variation seminorm is a good choice of \mathcal{R} . However, we point out that for the total variation seminorm in a Bayesian setting there exists no straightforward useful infinite dimensional limit (see [16]). Bayesian prior modelling in infinite dimensional settings is often considerably more involved.

If \mathbf{f} is a probability density, a frequent choice of the penalty functional is $\mathcal{R}(\mathbf{f}) = \text{KL}(\mathbf{f}, \mathbf{f}_0)$, which naturally enforces nonnegativity of the solution because of the logarithm. Alternatively, more general inequality constraints $N(\mathbf{f}) \leq 0$ for some function N can be incorporated into the penalty function by replacing \mathcal{R} by

$$\tilde{\mathcal{R}}(\mathbf{f}) = \begin{cases} \mathcal{R}(\mathbf{f}), & \text{if } N(\mathbf{f}) \leq 0 \\ \infty, & \text{else.} \end{cases}$$

5.2.1.1 Implementation

In this paragraph we will discuss several possibilities to compute the minimizer $\hat{\mathbf{f}}_\alpha$ of (5.9) for a linear forward operator denoted by $F = T$. In the case of quadratic Tikhonov regularization it follows from the first order optimality conditions that the Tikhonov functional J_α has the unique minimizer

$$\hat{\mathbf{f}}_\alpha = (T^*T + \alpha I)^{-1} (T^*\mathbf{g} + \alpha\mathbf{f}_0) \quad (5.10)$$

for all $\alpha > 0$. So in order to compute the regularized solution we have to solve the linear system of equations

$$\mathbf{A}\mathbf{f} = \mathbf{b} \quad \text{with } \mathbf{A} := T^*T + \alpha I \text{ and } \mathbf{b} = T^*\mathbf{g} + \alpha\mathbf{f}_0.$$

Solving this directly by for example Gauss-Jordan elimination requires $\mathcal{O}(n^3)$ operations and we have to store the full matrix A . In imaging applications n is typically the number of pixels or voxels, which can be so large that storing A is impossible. Therefore, we have to resort to iterative methods which access A only via matrix-vector products. Such matrix-vector products can often be implemented efficiently without setting up the matrix A , e.g. by the fast Fourier transform (FFT) or by solving a partial differential equation. As A is positive definite, the most common method for solving $\mathbf{A}\mathbf{f} = \mathbf{b}$ is the conjugate gradient (CG) method:

Algorithm 5.1 (CG iteration for solving $\mathbf{A}\mathbf{f} = \mathbf{b}$, $A > 0$)

Initialization. Choose initial guess $\mathbf{f}_0 \in \mathbb{R}^n$. Set $\mathbf{s}_0 = \mathbf{b} - \mathbf{A}\mathbf{f}_0$; $\mathbf{d}_0 = \mathbf{s}_0$.

General Step ($l = 0, 1, \dots$)

$$\begin{aligned}\gamma_l &= \frac{\|\mathbf{s}_l\|_2^2}{\langle \mathbf{d}_l, \mathbf{A}\mathbf{d}_l \rangle} \\ \mathbf{f}_{l+1} &= \mathbf{f}_l + \gamma_l \mathbf{d}_l \\ \mathbf{s}_{l+1} &= \mathbf{s}_l + \gamma_l \mathbf{A}\mathbf{d}_l \\ \beta_l &= \frac{\|\mathbf{s}_{l+1}\|_2^2}{\|\mathbf{s}_l\|_2^2} \\ \mathbf{d}_{l+1} &= \mathbf{s}_{l+1} + \beta_l \mathbf{d}_l.\end{aligned}$$

If $A = T^*T + \alpha I$ as in Tikhonov regularization, the stopping criterion $\mathbf{s} \neq 0$ may be replaced by

$$\|\mathbf{s}_l\| > \frac{\text{TOL}}{\alpha}$$

for some tolerance parameter $\text{TOL} > 0$. It can be shown that $\mathbf{s}_l = \mathbf{b} - \mathbf{A}\mathbf{f}_l$ for all l . As $\|A^{-1}\| \leq 1/\alpha$, this guarantees the error bound $\|\bar{\mathbf{f}} - \mathbf{f}_L\| \leq \text{TOL}$ to the exact minimum $\bar{\mathbf{f}} = A^{-1}b$ of the Tikhonov functional.

In the case of more general data fidelity terms $S_{\mathbf{g}}$ and penalty terms \mathcal{R} one can use a primal-dual algorithm suggested by Chambolle and Pock [3]. To formulate this algorithm, recall that for a functional $\mathcal{F} : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ the conjugate functional is given by

$$\mathcal{F}^*(\mathbf{s}) := \sup_{\mathbf{x} \in \mathbb{R}^n} [\langle \mathbf{s}, \mathbf{x} \rangle - \mathcal{F}(\mathbf{x})]. \quad (5.11)$$

If \mathcal{F} is convex and continuous, then $\mathcal{F}^{**} = \mathcal{F}$. For more information on this and other basic notions of convex analysis used in the following we refer, e.g. to [17]. The algorithm is based on the saddle point formulation

$$\min_{\mathbf{f} \in \mathbb{R}^n} \max_{\mathbf{z} \in \mathbb{R}^m} [\langle T\mathbf{f}, \mathbf{z} \rangle + \alpha \mathcal{R}(\mathbf{f}) - S_{\mathbf{g}}^*(\mathbf{z})] = \max_{\mathbf{z} \in \mathbb{R}^m} \min_{\mathbf{f} \in \mathbb{R}^n} [\langle T\mathbf{f}, \mathbf{z} \rangle + \alpha \mathcal{R}(\mathbf{f}) - S_{\mathbf{g}}^*(\mathbf{z})].$$

Note that an analytic computation of the maximum leads to the original problem (5.9) whereas a computation of the minimum leads to the dual problem

$$\max_{\mathbf{z} \in \mathbb{R}^m} - \left[S_{\mathbf{g}}^*(\mathbf{z}) + \alpha \mathcal{R}^* \left(-\frac{1}{\alpha} T^* \mathbf{z} \right) \right] \quad (5.12)$$

The algorithm requires the computation of so called *proximity operators* (see Chap. 6). For a functional $G : \mathbb{R}^n \rightarrow \mathbb{R}$ and a scalar $\lambda > 0$ the proximity operator $\text{prox}_{G,\lambda} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is defined by

$$\text{prox}_{G,\lambda} := (\mathbf{z}) = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \left[\frac{1}{2} \|\mathbf{z} - \mathbf{x}\|_2^2 + \lambda G(\mathbf{x}) \right].$$

For many popular choices of $S_{\mathbf{g}}$ and \mathcal{R} the proximity operator can either be calculated directly via a closed expression or there are efficient algorithms for their computation.

To give a simple example, for $G(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|_2^2$ one can calculate the proximity operator just by (5.10) to be $(1 + \lambda)^{-1}\mathbf{z}$.

One also needs to evaluate the proximity operator of the *Fenchel convex conjugate* G^* , which can be done by Moreau's identity

$$\text{prox}_{G^*,\lambda}(\mathbf{z}) = \mathbf{z} - \lambda \text{prox}_{G,1/\lambda}\left(\frac{1}{\lambda}\mathbf{z}\right).$$

Algorithm 5.2 (*Chambolle-Pock primal dual algorithm*)

Initialization. Choose $(\mathbf{f}_0, \mathbf{p}_0) \in \mathbb{R}^n \times \mathbb{R}^m$ and set $\tilde{\mathbf{f}}_0 = \mathbf{f}_0$.

General Step ($k = 0, 1, \dots$) Choose parameters $\tau_X^k, \tau_Y^k > 0, \theta_k \in [0, 1]$ and let

$$\begin{aligned}\mathbf{p}_{k+1} &= \text{prox}_{S_{\mathbf{g}},\tau_Y^k}\left(\mathbf{p}_k + \tau_Y^k T \mathbf{f}_k\right) \\ \mathbf{f}_{k+1} &= \text{prox}_{\alpha\mathcal{R},\tau_X^k}\left(\mathbf{f}_k - \tau_X^k T^* \mathbf{p}_k\right) \\ \tilde{\mathbf{f}}_{k+1} &= \theta_k (\mathbf{f}_{k+1} - \mathbf{f}_k).\end{aligned}$$

For constant parameters $\tau_X^k, \tau_Y^k > 0$ and $\theta_k = 1$ it has been shown [3] that \mathbf{f}_k converges to a solution of (5.9) and \mathbf{p}_k converges to a solution of the corresponding dual problem. Under certain assumptions on $S_{\mathbf{g}}$ and \mathcal{R} special choices for the parameters $\tau_X^k, \tau_Y^k > 0$ and θ_k will speed up the convergence.

To compute the minimizer of (5.9) with additional inequality constraints one can apply semismooth Newton methods for which we refer to [19].

5.2.2 Iterative Regularization

Whereas in the previous subsection we have discussed iterative methods to compute the minimum of a generalized Tikhonov functional, in the following we will discuss iterative methods for the solution of $F(\mathbf{f}) = \mathbf{g}$ without prior regularization. Typically the choice of the stopping index plays the role of the choice of the regularization parameter α . A motivation for the use of iterative regularization method is the fact that the Tikhonov functional is not convex in general for nonlinear operators. Therefore, it cannot be guaranteed that an approximation to the global minimum of the Tikhonov functional can be computed. For further information on iterative regularization methods we refer to the monograph [15].

5.2.2.1 Landweber Iteration

Landweber iteration can be derived as a method of steepest descent for the cost functional $J(\mathbf{f}) = \frac{1}{2}\|F(\mathbf{f}) - \mathbf{g}\|_2^2$. As the direction of steepest descent is given by the negative gradient $-J'(\mathbf{f}) = -F'(\mathbf{f})^*(F(\mathbf{f}) - \mathbf{g})$, this leads to the iteration formula

$$\mathbf{f}_{k+1} = \mathbf{f}_k - \mu F'(\mathbf{f}_k)^*(F(\mathbf{f}_k) - \mathbf{g}) \quad (5.13)$$

with a step size parameter μ . This parameter should be chosen such that $\mu \|F'(\mathbf{f})^* F'(\mathbf{f})\| < 1$ for all \mathbf{f} . Since a too small choice of μ slows down convergence considerably, it is advisable to compute the operator norm of $F'(\mathbf{f})^* F'(\mathbf{f})$ by a few iterations of the power method.

Under certain conditions on the operator F it has been shown in [7] in a Hilbert space setting that Landweber iteration with the discrepancy principle as stopping rule is a regularization method in a sense analogous to Definition 5.1, i.e. the worst case error tends to 0 with the noise level for a sufficiently good initial guess.

5.2.2.2 Regularized Newton Methods

Although Landweber iteration often makes good progress in the first few iterations, asymptotic convergence is very slow. Faster convergence may be expected from Newton-type methods, which solve a linear system of equations or some minimization problem using the first order Taylor approximation

$$F(\mathbf{f}) \approx F(\mathbf{f}_k) + F'(\mathbf{f}_k)(\mathbf{f} - \mathbf{f}_k). \quad (5.14)$$

around a current iterate \mathbf{f}_k . Plugging this approximation into the quadratic Tikhonov functional and using the last iterate as initial guess leads to the *Levenberg-Marquardt-algorithm*

$$\mathbf{f}_{k+1} = \operatorname{argmin}_{\mathbf{f} \in \mathbb{R}^n} \left[\frac{1}{2} \|F'(\mathbf{f}_k)(\mathbf{f} - \mathbf{f}_k) + F(\mathbf{f}_k) - \mathbf{g}\|_2^2 + \frac{\alpha_k}{2} \|\mathbf{f} - \mathbf{f}_k\|_2^2 \right].$$

For a convergence analysis we refer to [6]. The minimization problems can be solved efficiently by Algorithm 5.1 without the need to set up the full Jacobi matrices $F'(\mathbf{f}_k)$ in each step. Newton-type methods converge considerably faster than Landweber iteration. In fact the number of Landweber steps which is necessary to achieve an accuracy comparable to k Newton steps increases exponentially with k (cf. [4]). On the other hand, each iteration step is more expensive. Which of the two methods is more efficient may depend on the size of the noise level. Newton-type methods are typically favorable for small noise levels. Plugging the first order Taylor approximation (5.14) into the generalized Tikhonov functional (5.8) leads to the iteration formula

$$\mathbf{f}_{k+1} \in \operatorname{argmin}_{\mathbf{f}} \left[\mathcal{S}_{g^{\text{obs}}} (F(\mathbf{f}_k) + F'(\mathbf{f}_k)(\mathbf{f} - \mathbf{f}_k)) + \alpha \mathcal{R}(f) \right]. \quad (5.15)$$

If $\mathcal{S}_{g^{\text{obs}}}(\mathbf{g}) = \frac{1}{2} \|\mathbf{g} - g^{\text{obs}}\|_2^2$ and $\mathcal{R}(\mathbf{f}) = \frac{1}{2} \|\mathbf{f} - \mathbf{f}_0\|_2$, this leads to the commonly used *iteratively regularized Gauss-Newton method* where compared to the Levenberg-Marquardt method the penalty term is replaced by $\frac{\alpha_k}{2} \|\mathbf{f} - \mathbf{f}_0\|_2^2$. Note that if $\mathcal{S}_{g^{\text{obs}}}$ and \mathcal{R} are convex, a convex optimization problem has to be solved in each Newton step, which can be done, e.g., by Algorithm 5.2. For a convergence analysis of (5.15) including the case of Poisson data we refer to [12].

5.3 Error Estimates

5.3.1 General Error Bounds for Variational Regularization

Under the deterministic noise model (5.4) we are now looking for error bounds of the form

$$\|\hat{f}_\alpha - f^\dagger\|^2 \leq \phi(\delta, \alpha) \quad (5.16)$$

for the Tikhonov estimator \hat{f}_α . In a second step the right hand side may be minimized over α . As discussed in Sect. 5.1.4 such estimates can only be obtained under additional conditions on f^\dagger , which are called source conditions. There are several forms of such conditions. Nowadays, starting with [8] such conditions are often formulated in the form of variational inequalities. For the sake of simplicity we confine ourselves to the case of Hilbert spaces with quadratic functionals \mathcal{R} and $\mathcal{S}_{g^{\text{obs}}}$ here, although the concepts can be generalized with little additional effort to general convex \mathcal{R} and $\mathcal{S}_{g^{\text{obs}}}$. For a concave, monotonically increasing and continuous function $\psi : [0, \infty) \rightarrow [0, \infty)$ with $\psi(0) = 0$ we require that

$$\forall f \in \mathbb{X} \quad \frac{1}{4} \|f - f^\dagger\|^2 \leq \frac{1}{2} \|f\|^2 - \frac{1}{2} \|f^\dagger\|^2 + \psi(\|F(f) - F(f^\dagger)\|^2). \quad (5.17)$$

Such conditions are not easy to interpret at first sight, and we will come back to this in Sect. 5.3.2. However, they have been shown to be necessary for certain convergence rates of Tikhonov regularization and other regularization methods (see [11]), and sufficiency can be shown quite easily:

Theorem 5.3 *If f^\dagger fulfill (5.17), then the Tikhonov estimator \hat{f}_α in (5.8) satisfies the error estimate*

$$\frac{1}{4} \|\hat{f}_\alpha - f^\dagger\|^2 \leq \frac{\delta^2}{\alpha} + (-\psi)^* \left(-\frac{1}{4\alpha} \right) \quad (5.18)$$

where $\psi^*(t) := \sup_{s \geq 0} (st - \psi(t))$ denotes the conjugate function (see (5.11)).

Proof By definition of \hat{f}_α and our noise model we have

$$\frac{1}{2} \|F(\hat{f}_\alpha) - g^{\text{obs}}\|^2 + \frac{\alpha}{2} \|\hat{f}_\alpha\|^2 \leq \frac{1}{2} \|F(f^\dagger) - g^{\text{obs}}\|^2 + \frac{\alpha}{2} \|f^\dagger\|^2 \leq \frac{\delta^2}{2} + \frac{\alpha}{2} \|f^\dagger\|^2,$$

so together with our assumption (5.17) we find

$$\begin{aligned} \frac{1}{4} \|\hat{f}_\alpha - f^\dagger\|^2 &\leq \frac{1}{2} \|\hat{f}_\alpha\|^2 - \frac{1}{2} \|f^\dagger\|^2 + \psi(\|F(\hat{f}_\alpha) - F(f^\dagger)\|^2) \\ &\leq \frac{\delta^2}{2\alpha} - \frac{1}{2\alpha} \|F(\hat{f}_\alpha) - g^{\text{obs}}\|^2 + \psi(\|F(\hat{f}_\alpha) - F(f^\dagger)\|^2). \end{aligned}$$

By the parallelogram law we have for all $x, y, z \in \mathbb{X}$ that

$$\|x - y\|^2 \leq 2\|x - z\|^2 + 2\|z - y\|^2 - \|x + y - 2z\|^2 \leq 2\|x - z\|^2 + 2\|z - y\|^2.$$

Apply this with $x = F(\hat{f}_\alpha)$, $y = F(f^\dagger)$ and $z = g^{\text{obs}}$ to find

$$\frac{1}{4\alpha} \|F(\hat{f}_\alpha) - F(f^\dagger)\|^2 \leq \frac{\delta^2}{2\alpha} + \frac{1}{2\alpha} \|F(\hat{f}_\alpha) - g^{\text{obs}}\|^2,$$

so that finally we have

$$\begin{aligned} \frac{1}{4} \|\hat{f}_\alpha - f^\dagger\|^2 &\leq \frac{\delta^2}{\alpha} - \frac{1}{4\alpha} \|F(\hat{f}_\alpha) - F(f^\dagger)\|^2 + \psi \left(\|F(\hat{f}_\alpha) - F(f^\dagger)\|^2 \right) \\ &\leq \frac{\delta^2}{\alpha} + \sup_{s \geq 0} \left[\frac{-s}{4\alpha} + \psi(s) \right] = \frac{\delta^2}{\alpha} + (-\psi)^* \left(-\frac{1}{4\alpha} \right). \quad \square \end{aligned}$$

The two most commonly used types of functions ψ in the literature,

$$\psi_\nu(t) = t^{\nu/2} \quad \text{and} \quad \psi_p^{\log}(t) = (-\log t)^p (1 + o(1)), \quad \text{as } t \rightarrow 0$$

are referred to as *Hölder* and *logarithmic* source function, respectively. For these functions we obtain

$$\begin{aligned} (-\psi_\nu)^* \left(-\frac{1}{t} \right) &= ct^{\frac{\nu}{2-\nu}} \\ (-\psi_p^{\log})^* \left(-\frac{1}{t} \right) &= (-\log t)^p (1 + o(1)), \quad \text{as } t \rightarrow 0. \end{aligned}$$

Note that the two terms on the right hand side of (5.18) correspond to the error splitting (5.5). The following theorem gives an optimal choice of α balancing these two terms:

Theorem 5.4 *If under the assumptions of Theorem 5.3 ψ is differentiable, the infimum of the right hand side of (5.18) is attained at $\bar{\alpha}(\delta)$ if and only if*

$$\frac{1}{4\bar{\alpha}(\delta)} = \psi'(4\delta^2),$$

and

$$\frac{1}{4} \|\hat{f}_{\bar{\alpha}} - f^\dagger\|^2 \leq \psi(4\delta^2).$$

Proof Note from the definition of $(-\psi)^*$ that $(-\psi)^*(t^*) \geq tt^* + \psi(t)$ for all $t \geq 0$, $t^* \in \mathbb{R}$. Further equality holds true if and only if $t^* = -\psi'(t)$, as for this choice of t the concave function $\tilde{t} \mapsto \tilde{t}t^* + \psi(\tilde{t})$ attains its unique maximum at t and thus in particular $tt^* + \psi(t) \geq (-\psi)^*(t^*)$. Therefore we have with $t = 4\delta^2$ and $t^* = -\frac{1}{4\alpha}$ that

$$\frac{\delta^2}{\alpha} + (-\psi)^* \left(-\frac{1}{4\alpha} \right) \geq \psi(4\delta^2) \quad \text{for all } \alpha > 0$$

and $\frac{\delta^2}{\alpha} + (-\psi)^* \left(-\frac{1}{4\alpha} \right) = \psi(4\delta^2)$ if and only if $-\frac{1}{4\alpha} = -\psi'(4\delta^2)$. \square

It can be shown that the same type of error bound can be obtained by a version of the discrepancy principle (5.6) which does not require knowledge of the function ψ describing abstract smoothness of the unknown solution f^\dagger [2]. This is an advantage in practice, because such knowledge is often unrealistic.

5.3.2 Interpretation of Variational Source Conditions

5.3.2.1 Connection to Stability Estimates

Variational source conditions (5.17) are closely related to so called *stability estimates*. In fact if (5.17) holds true for all $f^\dagger \in \mathcal{K} \subset \mathbb{X}$, then all $f_1, f_2 \in \mathcal{K}$ satisfy the stability estimate

$$\frac{1}{4} \|f_1 - f_2\|^2 \leq \psi \left(\|F(f_1) - F(f_2)\|^2 \right),$$

with the same function ψ , since one of the terms $\pm (\|f_1\|^2 - \|f_2\|^2)$ will be non-positive. There exists a considerable literature on such stability estimates (see e.g. [1, 14]). However it is unclear if stability estimates also imply variational source conditions as two difficulties have to be overcome. Firstly the term $\|f\|^2 - \|f^\dagger\|^2$ might be negative and secondly one would have to extend the estimate from the set \mathcal{K} to the whole space \mathbb{X} .

5.3.2.2 General Strategy for the Verification of Variational Source Conditions

In general the rate at which the error of reconstruction methods converges to 0 as the noise level tends to 0 in inverse problems depends on two factors: The smoothness of the solution f^\dagger and the degree of ill-posedness of F . We will describe both in terms of a family of finite dimensional subspaces $V_n \subset \mathbb{X}$ or the corresponding orthogonal projections $P_n : \mathbb{X} \rightarrow V_n$. The smoothness of f^\dagger will be measured by how fast the best approximations $P_n f^\dagger$ in V_n converge to f^\dagger :

$$\|(I - P_n) f^\dagger\|_{\mathbb{X}} \leq \kappa_n \tag{5.19a}$$

$$\lim_{n \rightarrow \infty} \kappa_n = 0 \tag{5.19b}$$

Inequalities of this type are called Bernstein inequalities, and they are well studied for many types of subspaces V_n such as spaces of polynomials, trigonometric polyno-

mials, splines, or finite elements. We will illustrate this for the case of trigonometric polynomials below.

Concerning the degree of ill-posedness, recall that any linear mapping on a finite dimensional space is continuous. Therefore, a linear, injective operator T restricted to a finite dimensional space V_n has a continuous inverse $(T|_{V_n})^{-1}$ defined on $T(V_n)$. However, the norm of these operators will grow with n , and the rate of growth may be used to measure the degree of ill-posedness. In the nonlinear case we may look at Lipschitz constants σ_n such that $\|P_n f^\dagger - P_n f\| \leq \sigma_n \|F(P_n f^\dagger) - F(P_n f)\|$. However, to obtain optimal results it turns out that we need estimates of inner products of $P_n f^\dagger - P_n f$ with f^\dagger . Moreover, on the right hand side we have to deal with $F(f^\dagger) - F(f)$ rather than $F(P_n f^\dagger) - F(P_n f)$:

$$\langle P_n f^\dagger, f^\dagger - f \rangle \leq \sigma_n \|F(f^\dagger) - F(f)\|_{\mathbb{Y}} \quad (5.19c)$$

The growth rate of σ_n describes what we will call *local degree of ill-posedness* of F at f^\dagger .

Theorem 5.5 *Let \mathbb{X} and \mathbb{Y} be Hilbert spaces and suppose that there exists a sequence of projection operators $P_n: \mathbb{X} \rightarrow \mathbb{X}$ and sequences $(\kappa_n)_{n \in \mathbb{N}}$, $(\sigma_n)_{n \in \mathbb{N}}$ of positive numbers such that (5.19) holds true for all $n \in \mathbb{N}$.*

Then f^\dagger fulfills a variational source condition (5.17) with the concave, continuous, increasing function

$$\psi(\tau) := \inf_{n \in \mathbb{N}} [\sigma_n \sqrt{\tau} + \kappa_n^2], \quad (5.20)$$

which satisfies $\psi(0) = 0$.

Proof By straightforward computations we see that the variational source condition (5.17) has the equivalent form

$$\forall f \in \mathbb{X}: \quad \langle f^\dagger, f - f^\dagger \rangle \leq \frac{1}{4} \|f - f^\dagger\|^2 + \psi(\|F(f) - F(f^\dagger)\|^2).$$

Using (5.19a), (5.19c), and the Cauchy-Schwarz inequality we get for each $n \in \mathbb{N}$ that

$$\begin{aligned} \langle f^\dagger, f^\dagger - f \rangle &= \langle P_n f^\dagger, f^\dagger - f \rangle + \langle (I - P_n) f^\dagger, f^\dagger - f \rangle \\ &\leq \sigma_n \|F(f^\dagger) - F(f)\|_{\mathbb{Y}} + \kappa_n \|f^\dagger - f\|_{\mathbb{X}} \\ &\leq \sigma_n \|F(f^\dagger) - F(f)\|_{\mathbb{Y}} + \kappa_n^2 + \frac{1}{4} \|f^\dagger - f\|_{\mathbb{X}}^2 \\ &\leq \sigma_n \|F(f^\dagger) - F(f)\|_{\mathbb{Y}} + \kappa_n^2 + \frac{1}{4} \|f - f^\dagger\|^2. \end{aligned}$$

Taking the infimum over the right hand side with respect to $n \in \mathbb{N}$ yields (5.20) with $\tau = \|F(f^\dagger) - F(f)\|_{\mathbb{Y}}^2$. As ψ is defined by an infimum over concave and increasing functions, it is also increasing and concave. Moreover, (5.19b) implies $\psi(0) = 0$. \square

5.3.2.3 Example: Numerical Differentiation

Recall that the trigonometric monomials $e_n(x) := e^{2\pi nix}$ form an orthonormal basis $\{e_n : n \in \mathbb{Z} \setminus \{0\}\}$ of $L^2_\diamond([0, 1])$, i.e. every function $f \in L^2_\diamond([0, 1])$ can be expressed as $f(x) = \sum_{n \in \mathbb{Z} \setminus \{0\}} \widehat{f}(n) e_n(x)$ with $\widehat{f}(n) = \int_0^1 f(x) \overline{e_n(x)} dx$. Further note that from the definition (5.3) of the forward operator we get

$$F(e_n) = \frac{1}{2\pi in} e_n, \quad n \in \mathbb{Z} \setminus \{0\}.$$

and that the k th derivative of f has Fourier coefficients $\widehat{f^{(k)}}(n) = (2\pi in)^k \widehat{f}(n)$. Therefore, the norm

$$\|f\|_{H^s} := \left(\sum_{n \in \mathbb{Z} \setminus \{0\}} (2\pi n)^{2s} |\widehat{f}(n)|^2 \right)^{1/2}$$

(called *Sobolev norm* of order $s \geq 0$) fulfills $\|f\|_{H^k} = \|f^{(k)}\|_{L^2}$ for $k \in \mathbb{N}_0$, but it also allows to measure non-integer degrees of smoothness of f .

We choose P_n as the orthogonal projection

$$P_n f := \sum_{0 < |m| \leq n} \widehat{f}(m) e_m.$$

Suppose that the k th distributional derivative belongs to $L^2_\diamond([0, 1])$. Then

$$\begin{aligned} \|(I - P_n) f^\dagger\|_{L^2}^2 &= \sum_{|m| > n} \left| \widehat{f^\dagger}(m) \right|^2 = \sum_{|m| > n} (2\pi m)^{-2s} \left| \widehat{(f^\dagger)^{(s)}}(m) \right|^2 \\ &\leq (2\pi n)^{-2s} \|f^\dagger\|_{H^s}^2 \end{aligned}$$

which shows that (5.19a) and (5.19b) are satisfied with $\kappa_n = (2\pi n)^{-k}$. Moreover, we have that

$$\begin{aligned} \langle P_n f^\dagger, f^\dagger - f \rangle &= \sum_{0 < |n| \leq m} -2\pi in \widehat{f^\dagger}(n) \overline{\left(\frac{1}{2\pi in} (\widehat{f^\dagger}(n) - \widehat{f}(n)) \right)} \\ &\leq \left(\sum_{0 < |n| \leq m} (2\pi n)^2 |\widehat{f^\dagger}(n)|^2 \right)^{1/2} \|F(f^\dagger) - F(f)\| \\ &\leq (2\pi m)^{1-s} \|f^\dagger\|_{H^s} \|F(f^\dagger) - F(f)\|. \end{aligned}$$

for $s \in (0, 1]$, so (5.19c) is satisfied with $\sigma_n := (2\pi m)^{1-s}$. Choosing the parameter $m \approx \|f^\dagger\|_{H^s}^{1/(1+s)} \tau^{-1/2(1+s)}$ at which the infimum in (5.20) is attained approximately

we see that there exists a constant $C > 0$ such that the variational source condition (5.17) is satisfied with $\psi(t) = C \|f^\dagger\|_{H^s}^{2/(1+s)} t^{s/(1+s)}$, and from Theorem 5.4 we obtain the error bound

$$\|\hat{f}_\alpha - f^\dagger\|_{L^2} = \mathcal{O}\left(\delta^{\frac{s}{s+1}}\right). \tag{5.21}$$

It can be shown that this rate is optimal in the sense that there exists no reconstruction method R for which

$$\inf_R \sup \left\{ \|R(g^{\text{obs}}) - f^\dagger\| : f^\dagger \in H^s, \|F(f^\dagger) - g^{\text{obs}}\| \leq \delta \right\} = o\left(\delta^{\frac{s}{s+1}}\right).$$

5.3.2.4 Example: Fluorescence Microscopy

Similarly one can proceed for the example of fluorescence microscopy. As one has to work here with L^2 spaces rather than L^2_\circ spaces, the Sobolev norm is defined by $\|f\|_{H^s}^2 = \sum_{n \in \mathbb{Z}} (1 + n^2)^s |\widehat{f}(n)|^2$. Assuming that the convolution kernel is a -times smoothing ($a > 0$) in the sense that $\|F(f)\|_{H^a} \sim \|f\|_{H^0}$, which is equivalent to the existence of two constant $0 < c < C$ such that the Fourier transform of the convolution kernel k fulfills

$$c(1 + |\xi|^2)^{-a/2} \leq |\widehat{k}(\xi)| \leq C(1 + |\xi|^2)^{-a/2}$$

one can show that $\|f^\dagger\|_{H^s} < \infty$ for $s \in (0, a]$ implies a variational source condition with $\psi(t) \sim t^{\frac{s}{s+a}}$ and an error bound

$$\|\hat{f}_\alpha - f^\dagger\|_{L^2} = \mathcal{O}\left(\delta^{\frac{s}{s+a}}\right). \tag{5.22}$$

Again this estimate is optimal in the sense explained above.

5.3.2.5 Extensions

Variational source conditions with a given Hölder source function actually hold true on a slightly larger set. In the typical situation where the marker density of the investigated specimen is constant (or smooth) up to jumps, then it fulfills the same variational source condition with $s = 1/2$, although $f^\dagger \in H^s$ if and only if $s < 1/2$. The sets on which a variational source condition is satisfied can be characterized in terms of Besov spaces $B^s_{2,\infty}$, and bounded subsets of such spaces are also the largest sets on which a Hölder-type error bound like (5.21) and (5.22) are satisfied with uniform constants (see [11]).

In the case where the convolution kernel is infinitely smoothing, e.g. if the kernel is a Gaussian, then we cannot expect to get a variational source condition with a Hölder source function under Sobolev smoothness assumptions. Instead one obtains

logarithmic source functions ψ_p^{\log} as introduced above, which will again be optimal and lead to very slowly decaying error estimates as $\delta \searrow 0$ [11].

Note that the rates (5.21) and (5.22) are restricted to smoothness indices $s \in (0, 1]$ and $s \in (0, a]$, respectively. This restriction to low order rates is a well-known shortcoming of variational source conditions. Higher order rates can be obtained by imposing a variational source condition on the dual problem (5.12), which can again be verified by Theorem 5.5 (see [5, 18]).

With some small modifications the strategy in Theorem 5.5 can be extended to Banach space settings [9, 21] and nonlinear forward operators F , in particular those arising in inverse scattering problems [10, 20].

5.3.3 Error Bounds for Poisson Data

We already briefly discussed discrete versions of inverse problems with Poisson in Sect. 5.2.1. Such problems arise in many photonic imaging modalities such as fluorescence microscopy, coherent x-ray imaging, positron emission tomography, but also electron microscopy. In the following we briefly discuss a continuous setting for such problems.

We consider a forward operator $F: \mathbb{X} \rightarrow \mathbb{Y}$ that maps an unknown sample $f^\dagger \in \mathbb{X}$ a photon density $g^\dagger \in \mathbb{Y} = L^1(\mathbb{M})$ generated by f^\dagger on some measurement manifold $\mathbb{M} \subset \mathbb{R}^d$. The given data are modeled by a Poisson process

$$G_t = \sum_{k=1}^N \delta_{x_k},$$

with density tg^\dagger . Here $\{x_1, \dots, x_N\} \subset \mathbb{M}$ denote the positions of the detected photons and $t > 0$ can be interpreted as exposure time. Note that $t \sim \mathbf{E}(N)$, i.e. the exposure time is proportional to the number of detected photons.

Now to discuss error bounds we first need some notion of noise level. But what is the “noise” in our setting? Our data G_t do not belong to $\mathbb{Y} = L^1(\mathbb{M})$ and the “noise” is not additive. However, it follows from the properties of Poisson processes that

$$\mathbf{E} \left[\langle \frac{1}{t} G_t, h \rangle \right] = \int_{\mathbb{M}} h g^\dagger dx,$$

and the variance of $\langle \frac{1}{t} G_t, h \rangle$ is proportional to $\frac{1}{t}$. This suggests that $\frac{1}{\sqrt{t}}$ plays the role of the noise level. More precisely, it is possible to derive concentration inequalities

$$\mathbf{P} \left(d \left(\frac{1}{t} G_t, g^\dagger \right) > \frac{r + C}{\sqrt{t}} \right) \leq \exp(-cr)$$

where the distance function d is defined by a negative Besov norm (see [13, 22] for similar results).

As a reconstruction method we consider generalized Tikhonov regularization as in (5.8) with \mathcal{S}_G , given by a negative (quasi-)log-likelihood corresponding to the Poisson data. As discussed in Sect. 5.2 this amounts to taking the Kullback-Leibler distance as data fidelity term in the finite dimensional case, and in particular in the implementations of this method. (Sometimes a small shift is introduced in the Kullback-Leibler divergence to “regularize” this term.) By assuming the VSC (5.17) with $\psi(\|F(\mathbf{f}^\dagger) - F(\mathbf{f})\|^2)$ replaced by $\psi(\text{KL}(F(\mathbf{f}^\dagger), F(\mathbf{f})))$ and an optimal parameter choice $\bar{\alpha}$ one can then show the following error estimate in expectation

$$\mathbf{E} \left(\|\hat{f}_{\bar{\alpha}} - f^\dagger\|_{\mathbb{X}} \right) = \mathcal{O} \left(\psi \left(\frac{1}{\sqrt{t}} \right) \right), \quad t \rightarrow \infty.$$

References

1. Alessandrini, G.: Stable determination of conductivities by boundary measurements. *Appl. Anal.* **27**, 153–172 (1988)
2. Anzengruber, S.W., Hofmann, B., Mathé, P.: Regularization properties of the sequential discrepancy principle for Tikhonov regularization in Banach spaces. *Appl. Anal.* **93**(7), 1382–1400 (2014). <https://doi.org/10.1080/00036811.2013.833326>
3. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**(1), 120–145 (2011). <https://doi.org/10.1007/s10851-010-0251-1>
4. Deuffhard, P., Engl, H.W., Scherzer, O.: A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under finely invariant conditions. *Inverse Probl.* **14**(5), 1081–1106 (1998). <https://doi.org/10.1088/0266-5611/14/5/002>
5. Grasmair, M.: Variational inequalities and higher order convergence rates for Tikhonov regularisation on Banach spaces. *J. Inverse Ill-Posed Probl.* **21**(3), 379–394 (2013). <https://doi.org/10.1515/jip-2013-0002>
6. Hanke, M.: A regularizing Levenberg-Marquardt scheme, with applications to inverse ground-water filtration problems. *Inverse Probl.* **13**, 79–95 (1997)
7. Hanke, M., Neubauer, A., Scherzer, O.: A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.* **72**, 21–37 (1995)
8. Hofmann, B., Kaltenbacher, B., Pöschl, C., Scherzer, O.: A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Probl.* **23**(3), 987–1010 (2007). <https://doi.org/10.1088/0266-5611/23/3/009>
9. Hohage, T., Miller, P.: Optimal convergence rates for sparsity promoting wavelet-regularization in Besov spaces. Technical report. <https://arxiv.org/abs/1810.06316> (2018)
10. Hohage, T., Weidling, F.: Verification of a variational source condition for acoustic inverse medium scattering problems. *Inverse Probl.* **31**(7), 075,006 (2015). <https://doi.org/10.1088/0266-5611/31/7/075006>
11. Hohage, T., Weidling, F.: Characterizations of variational source conditions, converse results, and maxisets of spectral regularization methods. *SIAM J. Numer. Anal.* **55**(2), 598–620 (2017). <https://doi.org/10.1137/16M1067445>
12. Hohage, T., Werner, F.: Iteratively regularized Newton-type methods for general data misfit functionals and applications to Poisson data. *Numer. Math.* **123**, 745–779 (2013). <https://doi.org/10.1007/s00211-012-0499-z>
13. Hohage, T., Werner, F.: Inverse problems with poisson data: statistical regularization theory, applications and algorithms. *Inverse Probl.* **32**(9), 093,001 (2016). <https://doi.org/10.1088/0266-5611/32/9/093001>

14. Isaev, M.I., Novikov, R.G.: Effectivized Hölder-logarithmic stability estimates for the Gel'fand inverse problem. *Inverse Probl.* **30**(9), 095,006 (2014). <https://doi.org/10.1088/0266-5611/30/9/095006>
15. Kaltenbacher, B., Neubauer, A., Scherzer, O.: *Iterative Regularization Methods for Nonlinear ill-posed Problems*. Radon Series on Computational and Applied Mathematics. de Gruyter, Berlin (2008)
16. Lassas, M., Siltanen, S.: Can one use total variation prior for edge-preserving bayesian inversion? *Inverse Probl.* **20**(5), 1537–1563 (2004). <https://doi.org/10.1088/0266-5611/20/5/013>
17. Rockafellar, R.T.: *Convex Analysis*. Princeton University Press (2015)
18. Sprung, B., Hohage, T.: Higher order convergence rates for Bregman iterated variational regularization of inverse problems. *Numer. Math.* **141**, 215–252 (2019). <https://doi.org/10.1007/s00211-018-0987-x>
19. Ulbrich, M.: *Semismooth Newton Methods For Variational Inequalities And Constrained Optimization Problems In Function Spaces*, MOS-SIAM Series on Optimization, vol. 11. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2011). <https://doi.org/10.1137/1.9781611970692>
20. Weidling, F., Hohage, T.: Variational source conditions and stability estimates for inverse electromagnetic medium scattering problems. *Inverse Probl. Imaging* **11**(1), 203–220 (2017). <https://doi.org/10.3934/ipi.2017010>
21. Weidling, F., Sprung, B., Hohage, T.: Optimal convergence rates for Tikhonov regularization in Besov spaces. Technical report (2018). <https://arxiv.org/abs/1803.11019>
22. Werner, F., Hohage, T.: Convergence rates in expectation for Tikhonov-type regularization of inverse problems with Poisson data. *Inverse Probl.* **28**(10), 104,004 (2012). <https://doi.org/10.1088/0266-5611/28/10/104004>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

