Regularization of inverse problems with large noise

Herbert Egger
Center for Computational Engineering Science, RWTH Aachen, Germany
E-mail: herbert.egger@rwth-aachen.de

Abstract. Regularization of ill-posed problems is only possible if certain bounds on the data noise level are available. We consider here the case of large, possibly unbounded noise, and propose a class of modified regularization methods that are capable of dealing with that case. After some modification, these methods can be analyzed by standard regularization theory, and optimal convergence rates are obtained. An analysis in the spirit of regularization in Hilbert scales allows to relate the results obtained to other approaches dealing with large noise, and to clarify the influence of the relaxed assumptions regarding the noise on the convergence rates. Finally, the theoretical results are illustrated by examples and numerical tests are presented.

1. Introduction
In this paper we consider linear inverse problems of the form

\[ T x = y^\delta, \]  

where \( T : \mathcal{X} \to \mathcal{Y} \) is a linear bounded operator between Hilbert spaces \( \mathcal{X} \) and \( \mathcal{Y} \) and \( y^\delta \) are the available, possibly perturbed, data. To simplify the presentation we assume that \( T \) is injective, and that problem (1) has a solution for unperturbed data \( y \) denoted by \( x^\dagger \).

It is well-known that, if (1) is ill-posed (e.g., if \( T \) is compact with infinite dimensional range), then a reasonable solution can only be found by regularization methods, cf., e.g., [4]. An important ingredient for regularization is the availability of some bound on the data noise, e.g., in a deterministic setting,

\[ \| y - y^\delta \| \leq \delta \]  

is assumed to hold for some \( \delta > 0 \). In case of random noise, that we do not treat here, bounds on the variance of the data noise are usually required, cf. [8, 1] for details and further references. Instead of (2) we will assume here that the data noise can be bounded reasonably only in a weaker norm \( \| \cdot \|_s \), which we are going to specify in the next section, i.e., we assume that only

\[ \| y - y^\delta \|_s \leq \delta \]  

holds instead of (2). For illustration, one might think of \( \mathcal{Y} \) being \( L^2 \) and \( \| \cdot \|_s \) being the \( H^{-1} \)-norm, but in general, we will choose the norm \( \| \cdot \|_s \) in dependence on the structure of the noise.

Problems including large deterministic noise have previously been considered, e.g., in [6] under a condition \( \| T^*(y^\delta - y) \| \leq \delta \). For regularization of problems with random or stochastic noise we again refer to [8, 1] and the references therein. An essential assumption in both cases is that the
operator $T$ respectively its adjoint are sufficiently smoothing, i.e., such that the propagation of the data error can be controlled appropriately. Additionally, the discrepancy principle (which might be considered to be the natural parameter choice strategy for iterative regularization methods) is not applicable, and therefore a-priori stopping rules or Lepskii-type principles are frequently used for large deterministic or random noise.

In contrast to previous works, we will not need such an assumption on the smoothness of the operator $T$, since we take account of the large noise by modifying our regularization methods. Our approach then allows to use the discrepancy principle as a stopping criterion also for the case of large noise (3) and to derive optimal convergence rates along the lines of standard regularization theory. A further analysis of our approach in the framework of regularization in Hilbert scales, cf. [11, 13, 15], allows a comparison with other approaches previously mentioned. In particular, we can clarify the influence of the relaxed assumptions regarding the noise on convergence rates.

The outline of this article is as follows: After recalling some results about Hilbert scales and motivating our modified regularization methods in the following section, we present the main convergence analysis in Section 3. In this section, we also relate our results to standard results from regularization in Hilbert scales and the ones presented in [6] for large deterministic noise. For illustration of our results, we present some examples and numerical tests in Section 4.

2. Preliminaries
The aim of this section is to introduce some prerequisites that are needed for the formulation of our modified regularization methods and for their analysis in the next section.

For the solution of problem (1), we will consider modified versions of spectral regularization methods. To keep the notation as simple as possible, we formulate our results for continuous regularization methods only; the generalization to iterative methods should be obvious. For illustration, let us consider Tikhonov regularization

$$x_\alpha := (T^*T + \alpha I)^{-1}T^*y^\delta.$$  (4)

In order to render (4) a regularization method, an appropriate parameter choice rule relating $\alpha$ to the noise level $\delta$ (and possibly $y^\delta$) has to be supplied, e.g., one might utilize a discrepancy principle, cf. [9, 4],

$$\alpha := \sup \{ \alpha > 0 : \|Tx_\alpha - y^\delta\| \leq \tau \delta \},$$  (5)

for some $\tau > 1$. If the data noise satisfies (2) for some $\delta > 0$, and the signal-to-noise ratio is in any way reasonable, i.e., $\|y\| > (\tau + 1)\delta$, then $\alpha$ is well-defined. Moreover, optimal convergence rates can be shown under appropriate source conditions, i.e., one has

$$\|x_\alpha - x^\delta\| = O(\delta^{\frac{2\mu}{4\mu+1}}) \quad \text{if} \quad x^\delta \in \mathcal{R}((T^*T)^\mu),$$  (6)

for some $0 < \mu \leq 1/2$.

In the case of large noise (3), i.e., if the data perturbation $(y^\delta - y)$ cannot be reasonably bounded in the norm $\|\cdot\|$, the above result is not directly applicable: in particular, the residual $(Tx_\alpha - y^\delta) \notin \mathcal{Y}$ in general, and thus the discrepancy principle (5) does not yield a well-defined stopping index. Without further assumptions on the smoothness properties of $T$, the right hand side $T^*y^\delta$ of the (regularized) normal equations will in general not be an element of $\mathcal{X}$. Hence, (4) does in general not yield a solution $x_\alpha \in \mathcal{X}$ for any $\alpha > 0$ if $y^\delta \notin \mathcal{Y}$. We will give particular examples of that situation in Section 4.

For choosing a norm $\|\cdot\|_s$, in which the data noise can be bounded appropriately, we will utilize the framework of Hilbert scales, see [7, 4] for details.

**Definition 1** Let $\mathcal{Y}$ be a real, separable Hilbert space, and let $L$ be a densely defined, unbounded, selfadjoint, strictly positive operator in $\mathcal{Y}$. Let us define the space $\mathcal{Y}_s$ as the completion of
\[ M := \bigcap_{k=0}^{\infty} D(L^k) \text{ with respect to the norm } \|y\|_s := \|L^s y\|_Y. \] The collection \( \{Y_s\}_{s \in \mathbb{R}} \) is then called a Hilbert scale.

Obviously, for \( s \in \mathbb{R} \), the space \( Y_s \) is again a Hilbert space and \( Y_0 = \mathcal{Y} \). For illustration, we suggest to think of \( Y_s = H^s \) being a Sobolev space, and refer to [12] for a discussion when Sobolev spaces in fact form a Hilbert scale.

We will now formulate a class of modified regularization methods by altering the problem setup, i.e., by formally considering \( T \) as an operator from \( \mathcal{X} \) to \( Y_s \) for some \( s \in \mathbb{R} \). The adjoint \( T' \) of \( T \) with respect to these spaces then has the form \( T' = T^* L^{2s} \), where \( T^* \) denotes the adjoint with respect to the natural spaces \( \mathcal{X} \) and \( \mathcal{Y} \). To give an example, consider the modified version of Tikhonov regularization, which reads

\[ x^\delta_\alpha = (T^* L^{2s} T + \alpha I)^{-1} T^* L^{2s} y^\delta. \] (7)

Under the standard assumption \( \|y - y^\delta\| \leq \delta \) on the noise, and assuming that

\[ \|T^* y\| \leq \|y\|_{-a} \quad \text{for all } y \in \mathcal{Y} \] (8)

holds for some \( a > 0 \), optimal convergence rates

\[ \|x^\delta_\alpha - x^\dagger\| = O(\delta \pi^a + \pi^u) \] (9)

can be proven to hold for method (7) for \( 0 \leq s \leq a/2 \) under a source condition

\[ x^\dagger = (B^* B)^{\frac{u}{2(a-s)}} v, \quad B := L^s T, \quad v \in \mathcal{X}, \] cf. [2]. Here, we are however interested in the complementary situation \( s < 0 \) and only require a relaxed assumption (3) on the noise. Therefore, we have to follow a different kind of analysis. Note that for \( s < 0 \) the operator \( L^s \) is smoothing, thus \( L^s y^\delta \) can be considered as a smoothed version of the data and \( L^s T \) is a smoothed version of the operator.

Remark 1 The idea to consider the operator \( T \) in (1) not as operator from \( \mathcal{X} \) to \( \mathcal{Y} \) but with respect to other spaces was introduced by Natterer [11] under the name regularization in Hilbert scales. Originally, \( T \) was considered as an operator acting form \( \mathcal{X}_s \) to \( \mathcal{Y} \) for some \( s > 0 \) with the aim to overcome saturation effects of Tikhonov regularization. The approach also allowed to relate abstract source conditions in (6) to smoothness requirements in terms of Sobolev spaces. Regularization in Hilbert scales was then generalized to other regularization methods and nonlinear problems in [13, 4, 15]. More recently, it was shown [3] that similar ideas can be utilized to formulate a preconditioning strategy for the solution of ill-posed problems, which yields to a remarkable speed-up of iterative regularization methods.

3. Convergence and Convergence rates

In this section we outline how convergence rates for the modified regularization methods follow directly from corresponding results of standard regularization theory. Moreover, we compare our results to the ones obtained by regularization in Hilbert scales under the noise assumption (2), and we clarify how the assumption (3) effects the convergence rates. Finally, we also compare our results to the ones obtained in [6] for large noise.

Let us start by rewriting the inverse problem (1) in the following way

\[ Bx = z^\delta, \quad B := L^s T, \quad z^\delta := L^s y^\delta, \] (10)
where we consider $T$ now as an operator acting from $X$ to $Y$. For ease of presentation, let us assume that $B$ is properly scaled, i.e., $\|B\| \leq 1$. The modified method (7) can then be interpreted as standard Tikhonov regularization for the modified problem (10). Denoting $z := L^s y$, the bound on the noise (3) reads

$$\|z^\delta - z\| \leq \delta. \quad (11)$$

Thus we can apply standard regularization theory to problem (10); in particular, the modified discrepancy principle

$$\alpha_\ast = \sup \left\{ \alpha > 0 : \|T x_\alpha^\delta - y\|_s \leq \tau \delta \right\} = \sup \left\{ \alpha > 0 : \|B x_\alpha^\delta - z_\delta\| \leq \tau \delta \right\} \quad (12)$$

with some appropriate $\tau > 1$ can be used for parameter selection in spectral regularization methods of the general form

$$x_\alpha^\delta = g_\alpha(T^s L^2 T'T^s L^2 y^\delta = g_\alpha(B^s B)B^s z_\delta. \quad (13)$$

Here, $g_\alpha$ and $r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda)$ are appropriate functions satisfying the standard assumptions

$$\sup_{\lambda \in (0,1)} |g_\alpha(\lambda)| \leq C a^{-1} \quad (14)$$

$$\sup_{\lambda \in (0,1)} \lambda^\mu |r_\alpha(\lambda)| \leq c a^\mu \quad \text{for all } 0 \leq \mu \leq \mu_0, \quad (15)$$

for $\alpha \in (0, \alpha_0]$ with some $\alpha_0, \mu_0 > 0$. In a similar way, modified versions of truncated singular value decomposition and iterative methods can be constructed by applying the standard methods to the modified problem (10). The natural source conditions for studying convergence of the modified methods (13) then are

$$x^\dagger = (B^s B)^\nu v, \quad \text{for some } \nu > 0 \text{ and } v \in X, \quad (16)$$

and optimal convergence rates of the modified methods follow directly from standard regularization theory applied to (10).

**Theorem 1** Let $x_\alpha^\delta$ be defined by (13) with $g_\alpha$ satisfying (14), (15). Additionally, let $x^\dagger$ denote the true solution of $Tx = y$, and assume that $y^\delta$ satisfies the bound (3) for some $s \in \mathbb{R}$. If the source condition (16) holds for some $0 \leq \nu \leq \mu_0$, then the rates

$$\|x_\alpha^\delta - x^{\dagger}\| = O(\delta^{\frac{2\nu}{\nu + 1}}) \quad (17)$$

hold for the parameter choice $\alpha \sim \delta^{\frac{2}{\nu + 1}}$. If, alternatively, $\alpha$ is determined by the discrepancy principle (12), then the rates (17) hold for $0 < \nu \leq \mu_0 - 1/2$.

**Remark 2** Note that we do not impose any smoothness assumption on $T$ in the formulation of Theorem 1. However, since we consider $T$ as an operator acting from $X$ to $Y$, the convergence rates (17) now depend on the smoothness of $x^\dagger$ measured in terms of the operator $B = L^s T$.

In order to see how the modification of the regularization methods respectively the relaxed assumption on the noise affects the convergence rates, let us look at our problem in the framework of Hilbert scales.
Remark 3 Assume that there exists positive constants $m$, $M$ and $a$ such that
\begin{equation}
\frac{m}{\mu} \|y\|_{-\alpha} \leq \|T^*y\| \leq \frac{M}{\mu} \|y\|_{-\alpha},
\end{equation}
i.e., the operators $L^{-\alpha}$ and $T^*$ are (norm) equivalent ($L^{-\alpha} \sim T^*$). A similar assumption is usually made for the analysis of regularization in Hilbert scales, cf. [3, 11]. It now follows from (18) and the inequality of Heinz that
\[ L^s \sim (TT^*)^{-\frac{s}{2\alpha}} \]
for all $-a \leq s \leq a$, and consequently
\[ B^s = T^*L^s \sim T^*(TT^*)^{-\frac{s}{2\alpha}} \sim (T^*T)^{-\frac{s-a}{2\alpha}} \]
for $-a \leq s \leq a$.

Hence the source condition $x^\dagger = (B^*B)^\nu w$ is equivalent to $x^\dagger = (T^*T)^\mu v$ with $\mu = \nu \frac{a-s}{a}$ for $-a \leq s \leq a$ and $|v| \leq 1/2$ and for some $v \in X$ with $\|v\| \sim \|w\|$. On the other hand, we can define spaces $X_u := \mathcal{R}(T^*L^{a-u})$. By virtue of (18) we then have $(T^*T)^\mu \sim T^*L^{a-u}$ for $0 \leq \mu \leq 1$ and $u = 2a\mu$, and thus the following range identities hold:
\begin{equation}
X_u = \mathcal{R}((T^*T)^{\frac{u}{2\alpha}}) = \mathcal{R}((B^*B)^{\frac{u}{2\alpha}}) \quad \text{for} \quad 0 \leq u \leq a - s, \quad |s| \leq a.
\end{equation}
This allows us to interpret the source condition (16) in various ways and to relate the convergence rate (17) to the ones of standard regularization methods (6).

In the notation used in Hilbert scale regularization, the convergence rates of Theorem 1 then read
\begin{equation}
\|x^{\dagger}_a - x^\dagger\| = O(\delta^{\frac{u}{2\alpha}}) \quad \text{for} \quad x^\dagger \in X_u.
\end{equation}
In fact, (20) follows directly by (19) and $u = 2(a-s)\nu$ in (17). For $s < 0$, the rate (20) is smaller than the usual rate $O(\delta^{\frac{u}{2\alpha}})$ which would hold if the stronger noise level bound (2) was valid. Thus, the weaker bound on the noise level leads to a reduction of the convergence rates; the rates (17) are however optimal under the given assumptions, which is clear from Theorem 1 and converse results, cf. [4].

Let us now relate our results also to the treatment of large noise in [6], where a condition
\begin{equation}
\|T^*(y - y^\delta)\| \leq \delta
\end{equation}
is used as bound for the noise and standard regularization methods (4) are applied.

Remark 4 Under assumption (18), condition (21) is equivalent to (3) with $s = -a$. Using the range identities (19) we arrive at the rates
\begin{equation}
\|x^{\dagger}_a - x^\dagger\| = O(\delta^{\frac{u}{2\alpha}}) = O(\delta^{\frac{2a\mu}{2\alpha}}) \quad \text{for} \quad x^\dagger \in X_u = \mathcal{R}((T^*T)^\mu)
\end{equation}
with $u = 2a\mu$. In case $s = -a$ and under the assumption (18) this result coincides with the one of Theorem 1. However, if $s \neq -a$, then the noise is overestimated and the convergence rates are suboptimal. On the other hand, if $s < -a$, then standard methods like by Tikhonov regularization (4), are not even well defined, i.e., $x^\delta_a \notin X$ in general. Note that the condition (18) was only used to relate the results of Theorem 1 to the ones obtained for the other approaches, but the result of Theorem 1 holds without such a condition.

We will now illustrate with examples, that our approach allows to treat large noise in a very natural, flexible and efficient way.
4. Examples and numerical tests

Throughout our examples we consider a Hilbert scale generated by an operator $L$ defined by $L^2 u := -\Delta u + u$, with $\mathcal{D}(L^2) = \{ u \in H^2(\Omega) : \frac{\partial u}{\partial n} = 0 \} \subset L_2(\Omega)$. For $\Omega = (0,1)$ we have the simple representation

$$Lu = \sum_n \sqrt{n^2 + 1} u_n \cos(n\pi t),$$

where $u = \sum_n u_n \cos(n\pi t)$.

For illustration, we consider here the following noise model: Let $\phi_n$ be an orthonormal basis for $\mathcal{Y}$, e.g., $\phi_n = \sqrt{2} \cos(n\pi t)$ for $n \geq 1$ and $\phi_0 = 1$ for $\mathcal{Y} = L_2(0,1)$. Then we assume that the $n$-th Fourier coefficient $y_n$ of our data $y = \sum_n y_n \phi_n$ is disturbed by some $\eta_n \in [\delta, \delta]$. Setting $y^n = y + \eta$ we obtain that

$$||y^n - y||_s \leq C_s \delta, \quad \text{for } s < -1/2.$$

On the other hand, the noise is not bounded in $\mathcal{Y}$, for $s \geq -1/2$, in particular not for $s = 0$. We want to note that our results do not depend on the specific noise model, but only on the availability of a bound (3). For an account on frequently used statistical noise models, we refer to [1] and the references therein.

Let us now demonstrate in a first example that for mildly ill-posed problems, condition (21) might not be sufficient to bound the noise and thus standard regularization methods can be infeasible.

**Example 1** Let $\Omega = (0,1)$ and consider the fractional integral operator, cf. [14],

$$(Tx)(s) := \int_0^s (s - t)^{\beta - 1} x(t) dt$$

for some $0 < \beta < 1/2$. Then for $\epsilon > 0$ we have $\mathcal{R}(T) \subset H^{\beta - \epsilon}(0,1) = \mathcal{Y}_{\beta - \epsilon}$ and consequently $T^* : \mathcal{Y}_{\beta - \epsilon} \rightarrow L_2(0,1)$ is bounded.

Now fix $\beta = 1/4$ and consider the solution of $Tx = y^n$ with noise $\eta = y^n - y$ as above. Then $||T^*(y^n - y)||$ is unbounded, and standard Tikhonov regularization (4) is not even well defined as regularization method on $\mathcal{X}$, i.e., $x^n \notin \mathcal{X}$ in general.

Let us now consider the modified iteration (7) with $s = -1/2 - \epsilon$ for some $\epsilon > 0$. Then $\mathcal{R}(B^*) = \mathcal{R}(T^*L^s) \supset H^{\beta - s}_1 = H^{\beta + 1/2 + \epsilon}_s$, where $H^1_s := \{ y \in H^1(0,1) : y(0) = 0 \}$ and the fractional order spaces are defined by interpolation with $L^2(0,1)$. Thus, for $\beta = 1/4$ and $x^n \in H^{3/4}_{-\epsilon}$, the results of Theorem 1 apply with $\nu = 1/2$.

For a numerical test, we set $x^n(t) = H(t - 1/2)$ where $H$ denotes the Heaviside function. Then $x^n \in H^{1/2-\eta}$ for any $\eta > 0$. Thus the source condition (16) holds with $\nu = \frac{1-2\eta}{3+4\epsilon} \sim 1/3$ for $\epsilon, \eta \sim 0$, and we expect convergence at a rate of approximately $O(\delta^2)$, cf. the results of the numerical test reported in Table 1.

We want to emphasize once more that standard Tikhonov regularization (4) would not be applicable to this example, since $x^n$ may not even be well-defined in $\mathcal{X}$.

In the second example, we want to show that a bound (21) on the noise may yield suboptimal results.

**Example 2** Let $T$ be defined by $Tx = y$ with $y$ satisfying

$$-y'' = x \quad \text{on } (0,1), \quad y(0) = y(1) = 0.$$
Table 1. Convergence history for modified Tikhonov regularization 13 applied to Example 1. The numerically observed rates are $\alpha \sim \delta^{1.22}$ and $\|x^{\delta}_\alpha - x^\dagger\| = \delta^{0.41}$.

Table 2. Numerical results for Example (2); left: modified Tikhonov regularization (7), right: standard Tikhonov regularization with noise bound (3).

Let us remark that our results also apply to more general, e.g., logarithmic source conditions, cf. [5, 10]. Thus our framework applies also to exponentially ill-posed problems like inverse potential problems or the backwards heat equation.

5. Conclusion

The formulation of inverse problems highly depends on the choice of appropriate spaces respectively norms. In order to deal with large noise, we proposed to leave what might be considered a natural setting of the problem, and choose the image space out of a scale of spaces $\{Y_s\}_{s \in \mathbb{R}}$ such that the noise can be bounded in the norm of $Y_s$ appropriately. This approach results in modified regularization algorithms that, after altering the problem setting, can be analyzed by standard regularization theory. Using results from Hilbert scale regularization, the convergence rate results can be related to the ones expected under the standard assumption on the noise, i.e., the influence of the weaker bound on the noise on the convergence rates can be clarified. As illustrated by examples, this approach is more flexible than just assuming that $\|T^*(y^\delta - y)\|$ is bounded, and yields optimal convergence rates for a broad class of problems.

References

[1] N. Bissantz, T. Hohage, A. Munk, and F. Ruymgaart. Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, 45:2610–2636, 2007.

[2] H. Egger. Y-scale regularization. *SIAM J. Numer. Anal.*, 46:419–436, 2008.
[3] H. Egger and A. Neubauer. Preconditioning Landweber iteration in Hilbert scales. *Numer. Math.*, 101:643 – 662, 2005.

[4] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[5] T. Hohage. Logarithmic convergence rates of the iteratively regularized Gauss-Newton method for an inverse potential and an inverse scattering problem. *Inverse Problems*, 13:1279–1299, 1997.

[6] M. Yu Kozurin. Stable iteratively regularized gradient method for nonlinear irregular equations under large noise. *Inverse Problems*, 22:197–207, 2006.

[7] S. G. Krein and J. I. Petunin. Scales of Banach spaces. *Russian Math. Surveys*, 21:85–160, 1966.

[8] P. Mathé and S. Pereverzyev. Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods. *SIAM J. Numer. Anal.*, 38:1999–2021, 2001.

[9] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[10] M. T. Nair, E. Schock, and U. Tautenhahn. Morozov’s discrepancy principle under general source conditions. *J. Anal. Appl.*, 22:199–214, 2003.

[11] F. Natterer. Error bounds for Tikhonov regularization in Hilbert scales. *Appl. Anal.*, 18:29–37, 1984.

[12] A. Neubauer. When do Sobolev spaces form a Hilbert scale? *Proc. Amer. Math. Soc.*, 103:557–562, 1988.

[13] A. Neubauer. Tikhonov regularization of nonlinear ill-posed problems in Hilbert scales. *Appl. Anal.*, 46:59–72, 1992.

[14] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional Integrals and Derivatives: Theory and Applications*. Gordon and Breach, Switzerland, 1993.

[15] U. Tautenhahn. Error estimates for regularization methods in Hilbert scales. *SIAM J. Numer. Anal.*, 33:2120–2130, 1996.