Regularization methods for ill-posed problems in multiple Hilbert scales

Gisela L Mazzieri$^{1,2}$ and Ruben D Spies$^{1,3}$

$^{1}$ Instituto de Matemática Aplicada del Litoral, IMAL, CONICET-UNL, Guémes 3450, S3000GLN, Santa Fe, Argentina
$^{2}$ Departamento de Matemática, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Santa Fe, Argentina
$^{3}$ Departamento de Matemática, Facultad de Ingeniería Química, Universidad Nacional del Litoral, Santa Fe, Argentina

E-mail: glmazzieri@santafeconicet.gov.ar and rspies@santafe-conicet.gov.ar

Received 20 August 2011, in final form 5 March 2012
Published 5 April 2012
Online at stacks.iop.org/IP/28/055005

Abstract

Several convergence results in Hilbert scales under different source conditions are proved and orders of convergence and optimal orders of convergence are derived. Also, relations between those source conditions are proved. The concept of a multiple Hilbert scale on a product space is introduced, and regularization methods on these scales are defined, both for the case of a single observation and for the case of multiple observations. In the latter case, it is shown how vector-valued regularization functions in these multiple Hilbert scales can be used. In all cases, convergence is proved and orders and optimal orders of convergence are shown. Finally, some potential applications and open problems are discussed.

(Some figures may appear in colour only in the online journal)

1. Introduction

Quite often an inverse problem can be formulated as the need for determining $x$ in an equation of the form

$$Tx = y,$$

(1)

where $T$ is a linear bounded operator between two infinite-dimensional Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, the range of $T$, $\mathcal{R}(T)$, is non-closed and $y$ is the data, which are known, perhaps with a certain degree of error. It is well known that under these hypotheses, problem (1) is ill-posed in the sense of Hadamard [8]. The ill-posedness is reflected in the fact that $T^*$, the Moore–Penrose generalized inverse of $T$, is unbounded and therefore small errors or noise in the data $y$ can result in arbitrarily large errors in the corresponding approximated solutions (see [24, 23]), turning unstable all standard numerical approximation methods, making them unsuitable for most applications and inappropriate from any practical point of view. The so-called regularization
methods are mathematical tools designed to restore stability to the inversion process and consist essentially of parametric families of continuous linear operators approximating $T^\dagger$. The mathematical theory of regularization methods is very wide (a comprehensive treatise on the subject can be found in the book by Engl, Hanke and Neubauer [6]) and it is of great interest in a broad variety of applications in many areas such as medicine, physics, geology, geophysics, biology, image restoration and processing, etc.

There exist numerous ways of regularizing an ill-posed inverse problem. Among the most standard and traditional methods, we mention the Tikhonov–Phillips method [21, 27, 28], truncated singular value decomposition (TSVD), Showalter’s method, total variation regularization [1], etc. Among all regularization methods, probably the best known and most commonly and widely used is the Tikhonov–Phillips method, which was originally proposed by Tikhonov and Phillips in 1962 and 1963 (see [21, 27, 28]). Although this method can be formalized within a very general framework by means of spectral theory [6, 2], its widespread use is undoubtedly due to the fact that it can also be formulated in a very simple way as an optimization problem. In fact, the regularized solution of problem (1) obtained by applying the Tikhonov–Phillips method is also the minimizer $x_\alpha$ of the functional

$$J_\alpha(x) = \|Tx - y\|^2 + \alpha \|x\|^2,$$

where $\alpha$ is a positive constant known as the regularization parameter. The penalizing term $\alpha \|x\|^2$ in (2) not only induces stability but it also determines certain regularity properties of the approximating regularized solutions $x_\alpha$ and of the corresponding least-squares solution which they approximate as $\alpha \to 0^+$. Thus, for instance, it is well known that minimizers of (2) are always ‘smooth’ and, for $\alpha \to 0^+$, they approximate the least-squares solution of the minimum norm of (1), that is, $\lim_{\alpha \to 0^+} x_\alpha = T^\dagger y$. This method is more precisely known as the Tikhonov–Phillips method of order zero. Other penalizers in (2) can also be used. For instance, in his original articles [27, 28], Tikhonov considered the more general functional

$$J_{\alpha,L}(x) = \|Tx - y\|^2 + \alpha \|Lx\|^2,$$

where $L$ is an operator defined on a certain domain $\mathcal{D}(L) \subset \mathcal{X}$, into a Hilbert space $\mathcal{Z}$. Usually $L$ is a differential operator and hence it has a nontrivial nullspace. In spline smoothing problems, for instance (see [29]), $L$ is taken as the second derivative operator.

The use of (3) to regularize problem (1) automatically implies the a priori knowledge or assumption that the exact solution belongs to $\mathcal{D}(L)$. This approach gives rise to the theory of generalized inverses and regularization with seminorms (see for instance [6], chapter 8). The use of Hilbert scales becomes appropriate when there is no certainty that the exact solution is in fact an element of $\mathcal{D}(L)$.

The structure of this paper is as follows. In section 2, we briefly recall the theory of regularization methods in Hilbert scales. In section 3, we prove several convergence results in Hilbert scales under different source conditions and establish orders of convergence and optimal orders of convergence. Also relations between those source conditions are proved. In section 4, the concept of a multiple Hilbert scale on a product space is introduced, and regularization methods on these scales are defined, first for the case of a single observation and then for the case of multiple observations. In the latter case, it is shown how vector-valued regularization functions in these multiple Hilbert scales can be used. In all cases, convergence is proved and orders and optimal orders of convergence are shown. Finally, in section 5, potential applications and open issues are discussed.
2. Regularization in Hilbert scales

In this section, we will introduce the definition of a Hilbert scale and a few known results that will be needed later. All of them can be found in classical books and articles on the subject, such as [6] and [17].

Throughout this work, we will assume that \( L \) is a densely defined, unbounded, strictly positive self-adjoint operator on a Hilbert space \( \mathcal{X} \), so that \( L \) is closed and satisfies \( (Lx, y) = (x, Ly) \) for every \( x, y \in \mathcal{D}(L) \) and there exists a positive constant \( \gamma \) such that
\[
(Lx, x) \geq \gamma \|x\|^2 \quad \text{for every } x \in \mathcal{D}(L).
\]

(4)

Consider the set \( \mathcal{M} \) of all elements \( x \in \mathcal{X} \) for which all natural powers of \( L \) are defined, that is, \( \mathcal{M} = \bigcap_{s \in \mathbb{R}} \mathcal{D}(L^s) \). By using spectral theory, it can be easily shown that the fractional powers \( L^s \) are well defined over \( \mathcal{M} \) for every \( s \in \mathbb{R} \) and that
\[
\mathcal{M} = \bigcap_{s \in \mathbb{R}} \mathcal{D}(L^s)
\]
(5)

(see for instance [20] and [2]).

**Definition 2.1** (Hilbert scales). Let \( \mathcal{M} \) be defined as in (5). For every \( t \in \mathbb{R} \), we define
\[
\langle x, y \rangle_t = (L^t x, L^t y) \quad \text{for } x, y \in \mathcal{M}.
\]

(6)

It can be immediately seen that \( \langle \cdot, \cdot \rangle \) defines an inner product in \( \mathcal{M} \), which in turn induces a norm \( \|x\|_t = \|L^t x\| \). The Hilbert space \( \mathcal{X}_t \) is defined as the completion of \( \mathcal{M} \) with respect to this norm \( \| \cdot \|_t \). The family of spaces \( (\mathcal{X}_t)_{t \in \mathbb{R}} \) is called the Hilbert scale induced by \( L \) over \( \mathcal{X} \). The operator \( L \) is called a ‘generator’ of the Hilbert scale \( (\mathcal{X}_t)_{t \in \mathbb{R}} \).

**Remark 2.2.** Note that a Hilbert scale is a completely ordered (by set inclusion) parametric family of Hilbert spaces and if the operator \( L \) is bounded then \( \mathcal{X}_t = \mathcal{X} \) for every \( t \in \mathbb{R} \).

The following proposition constitutes one of the fundamental results for the treatment of inverse ill-posed problems in Hilbert scales.

**Proposition 2.3.** Let \( (\mathcal{X}_t)_{t \in \mathbb{R}} \) be the Hilbert scale induced by \( L \) over \( \mathcal{X} \). Then, the following is true.

(i) For every \( s, t \in \mathbb{R} \) such that \(-\infty < s < t < \infty\), the space \( \mathcal{X}_t \) is continuously and densely embedded in \( \mathcal{X}_s \).

(ii) Let \( s, t \in \mathbb{R} \). The operator \( L^{-s} \) defined on \( \mathcal{M} \) has a unique extension to \( \mathcal{X}_t \) which is an isomorphism (surjective isometry) from \( \mathcal{X}_s \) onto \( \mathcal{X}_t \). This extension, also denoted with \( L^{-s} \), is self-adjoint and strictly positive seen as an operator in \( \mathcal{X}_s \) with domain \( \mathcal{X}_t \), if \( t > s \). Also the identity \( L^{-s} = L^t \) is valid for the appropriate extensions. In particular \( (L^{-s})^{-1} = L^t \).

(iii) If \( s \geq 0 \), then \( \mathcal{X}_s = \mathcal{D}(L^s) \) and \( \mathcal{X}_{-s} = (\mathcal{X}_s)' \); that is, \( \mathcal{X}_{-s} \) is the topological dual of \( \mathcal{X}_s \) (with the topology induced by the norm in \( \mathcal{X} \)).

(iv) Let \( q, r, s \in \mathbb{R} \) be such that \(-\infty < q < r < s < \infty \) \( y x \in \mathcal{X}_q \). Then, the following interpolation inequality holds:
\[
\|x\|_r \leq \|x\|_{q}^\frac{r-q}{s-q} \|x\|_{s}^\frac{s-q}{r-q}.
\]

(7)
Proof. See [6], proposition 8.19. □

In the remainder of this section, we will state several results which will be of fundamental importance in the following sections. In all cases, we have included appropriate references where their proofs can be found.

**Theorem 2.4** (Heinz inequality). Let $A$ and $L$ be two linear, unbounded densely defined, strictly positive, self-adjoint operators on a Hilbert space $\mathcal{X}$ such that

$$D(A) \subset D(L)$$

and

$$\|Lx\| \leq \|Ax\| \quad \forall \ x \in D(A).$$

Then, for every $\nu \in [0, 1]$ there holds

$$D(A^\nu) \subset D(L^\nu)$$

and

$$\|L^\nu x\| \leq \|A^\nu x\| \quad \forall \ x \in D(A^\nu).$$

Proof. See [6], proposition 8.21, p 213 (see also [9] and [13]). □

**Remark 2.5.** It is important to point out here that the result of theorem 2.4 remains valid under slightly weaker hypotheses on the involved operators. More precisely, it can be shown that the result remains valid if the operators $A$ and $L$ satisfy conditions (8) and (9) and are self-adjoint and nonnegative instead of strictly positive.

**Lemma 2.6.** Let $T : \mathcal{X} \rightarrow \mathcal{Y}$ be a linear bounded operator between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ and $L$ a linear, densely defined, self-adjoint, unbounded and strictly positive operator on the space $\mathcal{X}$. Let $(\mathcal{X}_t)_{t \in \mathbb{R}}$ be the Hilbert scale induced by $L$ over $\mathcal{X}$. If there exist constants $0 < m \leq M < \infty$ and $a \in \mathbb{R}^+$ such that

$$m \|x\|_{-a} \leq \|Tx\| \leq M \|x\|_{-a} \quad \forall \ x \in \mathcal{X},$$

then $\mathcal{R}(T^+) = \mathcal{X}_0$ (that is, $\mathcal{R}(T^+) = D(L^\nu) = \mathcal{R}(L^{-a})$).

Proof. See [3]. □

**Remark 2.7.** Note that if (12) holds, then the operator $T$ is injective. Also note that (12) essentially says that the operator $T$ induces a norm on $\mathcal{X}$ which is equivalent to that inherited by $\mathcal{X}$ from the Hilbert scale of order $t = -a$, generated by the operator $L$ over $\mathcal{X}$. Hence, it is reasonable to think, in intuitive terms, that the degree of regularity induced by $T$ is equivalent to the degree of regularity induced by $L^{-a}$, and therefore the same happens with the degree of ill-posedness of their respective inverses.

**Theorem 2.8.** Let $T : \mathcal{X} \rightarrow \mathcal{Y}$ be a linear bounded operator between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ and $L$ a linear, densely defined, self-adjoint, unbounded and strictly positive operator on $\mathcal{X}$. Let $(\mathcal{X}_t)_{t \in \mathbb{R}}$ be the Hilbert scale induced by the operator $L$ over $\mathcal{X}$. Suppose that the operator $T$ satisfies (12) for some $a > 0$ and $0 < m \leq M < \infty$. Given $s > 0$, define $B = TL^{-s}$ where $L^{-s}$ is considered extended to all $\mathcal{X}$ in the sense of proposition 2.3 (ii). Then, for every $\nu \in [0, 1]$ we have that

$$m^n \|x\|_{-\nu(a+s)} \leq \|(B^*B)^{\frac{\nu}{s}}x\| \leq M^n \|x\|_{-\nu(a+s)}, \quad \forall \ x \in \mathcal{X},$$

(13)
\[ M^{-\gamma} \|x\|_{\gamma(\alpha,s)} \leq \|(B^*B)^{-\gamma} x\| \leq m^{-\gamma} \|x\|_{\gamma(\alpha,s)}, \quad \forall x \in D((B^*B)^{-\gamma}). \] (14)

Also

\[ \mathcal{R}((B^*B)^{\frac{\gamma}{2}}) = \mathcal{X}_{\gamma(\alpha+s)}. \] (15)

**Proof.** See [3] (see also [6], corollary 8.22, p 214). \( \square \)

**Remark 2.9.** If the operators \( L^{-1} y T^*T \) commute, then (15) also remains valid for \( \nu > 1 \). This result, which we will prove later on (theorem 3.10), will be of fundamental importance in the extension of some results on convergence of regularization methods in Hilbert scales, which will be presented in section 3.

The inequalities in (13) can be interpreted in a similar way as done for (12) in remark 2.7. In fact, taking as ‘unit of regularity’ the degree induced by the operator \( L^{-1} \), the respective degrees of regularity induced by \( L^{-s} \) and \( T \) are \( s \) and \( a \), respectively. Hence, the degree induced by \( B = TL^{-s} \) is \( a + s \), the degree induced by \( B^*B \) is \( 2(a + s) \) and, therefore, the degree of regularity induced by \( (B^*B)^{\gamma} \) is \( \frac{\gamma}{2}(a + s) = \nu(a + s) \).

The idea of using Hilbert scales for regularizing inverse ill-posed problems was first introduced by Natterer in 1984 [17] for the special case of the classical Tikhonov–Phillips method. In his work, Natterer regularized the problem \( Tx = y \) by minimizing the functional

\[ \|Tx - y\|^2 + \alpha \|x\|^2_\nu, \] (16)

over the space \( \mathcal{X}_\nu \), where \( \| \cdot \|_\nu \) denotes the corresponding norm in the Hilbert scale (see definition 2.1) and \( \|y\|^2 - y \| \leq \delta \).

In certain cases, it is possible that a value of \( s_0 > 0 \) be known for which we are absolutely sure that the exact solution \( x_0 \in \mathcal{X}_{s_0} \), where \( (\mathcal{X}_\gamma)_{\gamma \in \mathbb{R}} \) is the Hilbert scale induced by the operator \( L \) over \( \mathcal{X} \). In such cases, it is possible to proceed with regularization of the problem \( Tx = y \) by means of the traditional methods, by replacing the Hilbert space \( \mathcal{X} \) by \( \mathcal{X}_{s_0} \) and obviously \( T \) by its restriction to \( D(L^{s_0}) \). In other cases, however, it is possible that such a value of \( s_0 \) be not exactly known, although it could be reasonable to assume the existence of some \( u > 0 \) for which

\[ x^\dagger \in \mathcal{X}_u \] (17)

(although the exact value of \( u \) be unknown). It is precisely in this case in which Hilbert scales provide a solid mathematical framework for the development of convergent regularization methods which allow us to take advantage, in a optimal and ‘adaptive’ way, of the source condition (17) in order to obtain the best possible convergence speed, even though \( u \) is unknown.

The first result about convergence on Hilbert scales is due to Natterer [17] and it is presented in the next theorem.

**Theorem 2.10.** Let \( T \in L(\mathcal{X}, \mathcal{Y}) \) with \( \mathcal{X} \) and \( \mathcal{Y} \) Hilbert spaces, \( T^\dagger \) the Moore–Penrose generalized inverse of \( T \), \( L : D(L) \subset \mathcal{X} \rightarrow \mathcal{X} \) a linear, densely defined, self-adjoint, unbounded operator with \( L \geq \gamma \) for some \( \gamma > 0 \) and \( (\mathcal{X}_\gamma)_{\gamma \in \mathbb{R}} \) the Hilbert scale induced by \( L \) over \( \mathcal{X} \). Suppose also that condition (12) holds. Let \( s \geq 0 \) and \( B = TL^{-s} \), as in theorem 2.8. Let \( g_\alpha : [0, \|B\|^2] \rightarrow \mathbb{R} \), \( \alpha > 0 \), be a family of piecewise continuous functions and \( \rho_\alpha(\lambda) = 1 - \lambda g_\alpha(\lambda) \). Suppose also that \( \{g_\alpha\} \) satisfies the following conditions:

\[ C1: \forall \lambda \in (0, \|B\|^2], \text{ we have that } \lim_{\alpha \rightarrow 0^+} g_\alpha(\lambda) = \frac{1}{\lambda}, \] (18)
\[ \exists \hat{c} > 0 \text{ such that } \forall \lambda \in (0, \|B\|^2] \text{ and } \forall \alpha > 0 \text{ there holds } |g_\alpha(\lambda)| \leq \hat{c} \alpha^{-1}, \]

(19)

C3: \[ \exists \mu_0 \geq 1 \text{ such that if } \mu \in [0, \mu_0] \text{ then } \lambda^\mu |r_\alpha(\lambda)| \leq c_\mu \alpha^\mu \forall \lambda \in (0, \|B\|^2], \]

(20)

where \( c_\mu \) is a positive constant (here \( \mu_0 \) is the ‘qualification’ of the method).

For \( y \in D(T^\dagger) \), \( y^\delta \in Y \) with \( \|y - y^\delta\| \leq \delta \), we define the regularized solution of the problem

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

by

\[ x^\delta_\alpha = R_\alpha y^\delta = L^{-1} g_\alpha(B^*B)B^*y^\delta. \]

(21)

Suppose that \( x^\dagger = T^\dagger y \in X_u \) for some \( u \in [0, a + 2s] \) and that the regularization parameter \( \alpha \) is chosen as

\[ \alpha = c \left( \frac{\delta}{\|x^\dagger\|_u} \right)^{\frac{a+2s}{2+a}}. \]

(22)

where \( c \) is a positive constant and \( a \) is the constant in (12). Then, there exists a constant \( C \) (which depends on \( a \) and \( s \) but not on \( u \)) such that the following estimate for the total error holds:

\[ \|x^\delta_u - x^\dagger\|_u \leq C \|x^\dagger\|_u^{\frac{a+2s}{2+a}} \delta^{\frac{a}{2+a}}. \]

(23)

Proof. See [6], theorem 8.23.

In figure 1, the relation between the values of the parameters \( s \) and \( u \) of theorem 2.10 is schematized. Observe that the largest possible value for \( s \) is \( \frac{u-a}{2} \). The arrows indicate that the space \( X_u \) may or may not be contained in \( X^\delta_u \). The dashed curve represents the space \( X^\delta_u \) indicating that the parameter \( u \) is unknown.

Remark 2.11. It is very important to point out the ‘adaptivity’ of the order of convergence in theorem 2.10. In fact, note that although the regularized solutions \( x^\delta_u \) defined in (21) do not depend on the degree of regularity \( u \) of \( x^\dagger \) (it would be inappropriate if it depends on \( u \) since \( u \) is usually unknown), the order of convergence obtained does depend on \( u \). This order improves as \( u \) increases and it becomes asymptotically optimal in \( u \). Also observe that in order to ensure the order of convergence in (23) it is necessary to choose \( s \) (note that \( R_\alpha \) depends on \( s \)) such that \( u \leq a + 2s \). Since it is possible that \( u \) be unknown, it may happen that we may not be completely sure of the validity of such constraint. If this constraint was false, i.e. if \( u > a + 2s \), an order of convergence \( O(\delta^{\frac{a+2s}{2+a}}) \) cannot be guaranteed for the total error. However, since \( X_u \subset X^\delta_u \forall u \geq \eta \), in such circumstances we will still obtain at least convergence of the order \( O(\delta^{\frac{a+2s}{2+a}}) \) = \( O(\delta^{\frac{a+2s}{2+a}}) \). Thus, not choosing \( s \) sufficiently large will not result in lack of convergence but only in a worse (non-optimal) order of convergence.
3. Preliminary convergence results in Hilbert scales

In the next theorem, which extends the results of theorem 2.10, we will show that convergence can be obtained when the parameter choice rule $\alpha$ is chosen in the form $\alpha = c \delta^s$ for all values of $\delta$ in a certain interval, and not only for $\delta = \frac{2^a (a+1)}{a^2}$, corresponding to the choice in (22). We will prove however that for this choice of $\delta$ the order of convergence is optimal.

Theorem 3.1. Let $X, Y, T, T^\dagger, L, (X_t)_{t \in \mathbb{R}}$, $s \geq 0$, $a > 0$, $B = TL^{-s}$. $g_\alpha$, $r_\alpha$, $R_\alpha = L^{-s}g_\alpha(B^*B)^s$, $y \in D(T^\dagger)$, $\|y - y^0\| \leq \delta$, $u \in [0, a + 2s]$, $x^\dagger = T^\dagger y \in X_u$.

(i) moreover, $\|x^\dagger - x\| = O(\delta^s)$ where $\delta = \min \left\{1 - \frac{a}{2(a+1)}, \frac{a}{2(a+1)} \right\} > 0$.

(ii) the order of convergence for the total error is optimal when $\delta$ is chosen as $\delta = \frac{2^a (a+1)}{a^2}$, in which case $\|x^\dagger - x\| = O(\delta^\frac{a}{2a+2s})$.

Proof. First note that from conditions (19) and (20), it follows immediately that there exists a constant $k > 0$ such that

$$\lambda^\beta |g_\alpha(\lambda)| \leq k \alpha^s \beta^{1-a}, \quad \forall \alpha \in [0, 1], \forall \beta > 0 \quad \text{and} \quad \forall \lambda \in (0, [B]^2)$$

(24) (moreover we can take $k = \max\{1 + c_\alpha, \bar{c}\}$ where $c_\alpha$ is the constant $c_{1,\alpha}$ in (20) corresponding to $\mu = 0$).

We will now proceed to estimate the error due to noise in the data and the regularization error, separately. Without loss of generality we will suppose that $y \in \mathcal{R}(T)$ (otherwise we replace $y$ by $Qy$ where $Q : Y^\dagger \rightarrow \mathcal{R}(T)$; recall that $y \in D(T^\dagger)$ and $T^\dagger y = T^\dagger Qy$).

For the error due to noise, we have

$$\|x^\dagger - x\| = \|R_\alpha(y^\dagger - y)\|
= \|L^{-s}g_\alpha(B^*B)^s(y^\dagger - y)\|
= \|g_\alpha(B^*B)^s(y^\dagger - y)\|_s
\leq m^{-\frac{s}{a+s}} \|g_\alpha(B^*B)^s(y^\dagger - y)\|_s$$

(by (13) with $v = \frac{s}{a+s}$)

$$= m^{-\frac{s}{a+s}} \|g_\alpha(B^*B)^{-1/2}g_\alpha(B^*B)^s(y^\dagger - y)\|
= m^{-\frac{s}{a+s}} \|g_\alpha(B^*B)^{-1/2}g_\alpha(B^*B)^s(y^\dagger - y)\|
\leq m^{-\frac{s}{a+s}} \|g_\alpha(B^*B)^{-1/2}g_\alpha(B^*B)^s(y^\dagger - y)\|
= m^{-\frac{s}{a+s}} \|g_\alpha(B^*B)^{-1/2}g_\alpha(B^*B)^s(y^\dagger - y)\|
\leq C_1 \delta^\frac{a}{2a+2s} \|g_\alpha(B^*B)^{-1/2}g_\alpha(B^*B)^s(y^\dagger - y)\|$$

(24) with $\beta = \frac{a + 2s}{2a+2s}$.

where $C_1 = k m^{-\frac{s}{a+s}}$. Therefore,

$$\|x^\dagger - x\| \leq C_1 \delta^\frac{a}{2a+2s}.$$

(25)

At this point it is timely to observe that the estimate for the error due to noise in (25) is independent of the degree of regularity $u$ of the solution $x^\dagger$.

Next we proceed to estimate the regularization error $\|x_0 - x^\dagger\|$. Note in the first place that from proposition 2.3 (ii) (with $t = u$ and $s = u - s$), it follows that $L^s(a-s) = L^s$ has a unique extension to $X_u$, which is an isomorphism from $X_u$ onto $X_{u-s}$. It is important to point out here that it is precisely this property of the fractional powers of the operator $L$ on the Hilbert scale induced by itself, what will allow us, in the end, to arrive at the adaptive convergence order
that we want to prove. More precisely, note that whatever the value of $u$ (perhaps unknown), $L'$ always possesses a unique extension to $\mathcal{X}_u$. This extension, also denoted by $L'$, regarded as an operator on $\mathcal{X}_{u-s}$ with domain $\mathcal{X}_u$, is self-adjoint and strictly positive if $u > u - s$. Then, since $x^\dagger \in \mathcal{X}_u$, it follows that
\[ L'x^\dagger \in \mathcal{X}_{u-s}. \tag{26} \]

On the other hand, if $u \geq s$, from theorem 2.8 with $v \doteq \frac{u-s}{a+s}$ it follows that
\[ \mathcal{X}_{u-s} = \mathcal{R}((B^*B)^{\frac{s}{a+s}}). \tag{27} \]

From (26) and (27), it follows that there exists $v \in \mathcal{X}$ such that
\[ L'x^\dagger = (B^*B)^{\frac{s}{a+s}} v. \tag{28} \]

If $u < s$, then (28) holds with $v \doteq (B^*B)^{\frac{s}{a+s}} L'x^\dagger$.

Then,
\[
\|x_u - x^\dagger\| = \|R_u y - x^\dagger\| \\
= \|L^{-\gamma} g_u (B^*B) B'y - x^\dagger\| \\
= \|L^{-\gamma} g_u (B^*B) B' B'L'x^\dagger - x^\dagger\| \\
= \|L^{-\gamma} [g_u (B^*B) B'B' - I] L'x^\dagger\| \\
= \|L^{-\gamma} r_u (B^*B) B'L'x^\dagger\| \\
= \|L^{-\gamma} r_u (B^*B) (B^*B)^{\frac{s}{a+s}} v\| \quad \text{(by (28))} \\
= \|r_u (B^*B) (B^*B)^{\frac{s}{a+s}} v\|_{-s} \\
= \|\frac{s}{a+s} r_u (B^*B)\|_{-s} \quad \text{(by (13) with } v \doteq \frac{s}{a+s}) \]
\[
\leq m^{-\frac{s}{a+s}} \|\frac{s}{a+s} r_u (B^*B)\|_{-s} \quad \text{(by (20) with } \mu = \frac{u}{2(a+s)}) \\
\leq m^{-\frac{s}{a+s}} c_{\mu} x_{\mu} \|v\| \\
= \|\frac{s}{a+s} r_u (B^*B)\|_{-s} \quad \text{(by (28))} \\
\leq \|\frac{s}{a+s} c_{\mu} M\|_{-s} \|L'x^\dagger\|_{a-s} \\
= \|\frac{s}{a+s} c_{\mu} M\|_{-s} \|x^\dagger\|_a. \quad \text{(by (13) with } v \doteq \frac{s}{a+s}) \]

Hence, there exists $C_2 \doteq m^{-\frac{s}{a+s}} c_{\mu} M\|_{-s}$ such that
\[
\|x_u - x^\dagger\| \leq C_2\|x^\dagger\|_{a}\|u\|_{a}\|x^\dagger\|_a. \tag{29} \]

Note that this estimate for the regularization error does depend on the degree of regularity $u$ of $x^\dagger$ and is relevant only for the case $u > 0$.

Finally, from (25) and (29) it follows that
\[
\|x_u^\dagger - x^\dagger\| \leq \|x_u^\dagger - u\| + \|u - x^\dagger\| \\
\leq C_1 b\|\delta\|^\frac{s}{a+s} + C_2\|x^\dagger\|^\frac{s}{a+s} \|x^\dagger\|_u \\
= C_1 c\|\delta\|^\frac{s}{a+s} + C_2\|x^\dagger\|^\frac{s}{a+s} \|x^\dagger\|_u \\
= C_1 c\|\delta\|^\frac{s}{a+s} + C_2\|x^\dagger\|_u\|\delta\|^\frac{s}{a+s} \\
= \mathcal{O}(\delta^\sigma), \tag{30} \]

where $\sigma = \min\{1 - \frac{a}{2(a+s)}, \frac{m}{a+s}\}$. This proves (i) and (ii).
To prove \((iii)\), note that by virtue of \((30)\) it follows that the order of convergence is optimal when \(\varepsilon\) is chosen such that

\[
1 - \frac{ae}{2(a + s)} = \frac{\varepsilon u}{2(a + s)}.
\]

that is, for \(\varepsilon = \frac{2(a + s)}{a + s} - \frac{1}{a + s}\), in which case \(\sigma = \frac{u}{\varepsilon u}\). Finally, it is important to note here that this optimal order of convergence depends on \(a\) and \(u\) (that is, on \(L\), \(T\) and \(x^d\)) but it does not depend on the choice of \(s\). \(\square\)

In the next theorem, we will prove that with the same parameter choice rule as in \((22)\), it is possible to obtain a better order of convergence in a weaker norm or convergence in a stronger norm with a worse order.

**Theorem 3.2.** Let \(\mathcal{X}, \mathcal{V}, T, T^\dagger, L, (\mathcal{X}_c)_{c \in \mathbb{R}}\), \(s \geq 0\), \(a > 0\), \(B = TL^{-1}\), \(g_a, r_a, R_a, R_a = L^{-1}g_a(B^*B)B^*, y \in D(T^\dagger), y^\delta \in \mathcal{V}, \|y - y^\delta\| \leq \delta\), \(u \in [0, a + 2s]\), \(x^d = T^\dagger y \in \mathcal{X}_u\).

Theorem 3.2. Suppose that the parameter choice rule \(\alpha\) is chosen as in \((22)\), that is,

\[
\alpha = c \left( \frac{\delta}{\|x^d\|_u} \right)^{\frac{2(a + s)}{a + s}}.
\]

where \(c > 0\). Then, for every \(r \in [-a, \min\{u, s\}]\) there holds

\[
\|x_u^d - x_u\|_r \leq C \|x^d\|_u^{\frac{2(a + s)}{a + s}} \delta^{\frac{a + s}{a + s}},
\]

where \(C\) is a constant depending on \(a\), \(s\) and \(r\) but not on \(u\) nor on \(x^d\).

**Proof.** First note that due to the restriction on \(r\), we have that \(x^d, x_u, x^d_u\) are all in \(\mathcal{X}_u\). Just like in the previous theorem, without loss of generality we will assume that \(y \in R(T)\).

For the error due to noise, we have the following estimate:

\[
\|x_u^d - x_u\|_r = \|R_u(y^\delta - y)\|_r
= \|L^{-1}g_a(B^*B)B^*(y^\delta - y)\|_r
= \|g_a(B^*B)B^*(y^\delta - y)\|_{r=s}
\leq m^\delta \|B^*B\|^{\frac{2(a + s)}{a + s}} g_a(B^*B)B^*(y^\delta - y) \tag{by (13) with \(v = \frac{s - r}{a + s}\)}
\leq m^\delta \|B^*B\|^{\frac{2(a + s)}{a + s}} g_a(B^*B)B^*(y^\delta - y) \tag{by (24) with \(\beta = \frac{a + 2s - r}{2(a + s)}\)}
\leq m^\delta \|B^*B\|^{\frac{2(a + s)}{a + s}} y^\delta \tag{by (31)}
\leq C_1 \|x^d\|_u^{\frac{2(a + s)}{a + s}} \delta^{\frac{a + s}{a + s}},
\]

where \(C_1 = m^\delta \|B^*B\|^{\frac{2(a + s)}{a + s}}\). Thus,

\[
\|x_u^d - x_u\|_r \leq C_1 \|x^d\|_u^{\frac{2(a + s)}{a + s}} \delta^{\frac{a + s}{a + s}}.
\tag{33}
\]
For the regularization error, note that
\[
\| x_\alpha - x^1 \|_r = \| R_\alpha y - x^1 \|_r
\]
\[
= \| L^{-s} g_\alpha (B^* B) B^* y - x^1 \|_r
\]
\[
= \| L^{-s} g_\alpha (B^* B) B^* BL^s x^1 - x^1 \|_r
\]
(because \( B^* y = B^* BL^s x^1 \))
\[
= \| L^{-s} [g_\alpha (B^* B) B^* B - I] L^s x^1 \|_r
\]
\[
= \| L^{-s} r_\alpha (B^* B) L^s x^1 \|_r
\]
\[
= \| L^{-s} r_\alpha (B^* B) (B^* B)^* \hat{v} \|_r
\]
(by (28))
\[
= \| (B^* B)^* \hat{v} r_\alpha (B^* B) \|_r
\]
\[
= \| (B^* B) \hat{v} r_\alpha (B^* B) \|_{r-x}
\]
\[
\leq m_{\hat{v}} \| (B^* B)^* \hat{v} r_\alpha (B^* B) \|_{r-x}
\]
\[
= m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x}
\]
\[
\leq m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x} \leq (13) \text{ with } v = \frac{s - r}{a+s}
\]
\[
= m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x} \leq (13) \text{ with } \bar{\mu} = \frac{a - r}{2(a+s)} \cdot 0 \leq \bar{\mu} \leq 1
\]
\[
= m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x} \leq (13) \text{ with } v = \frac{s - u}{a+s}
\]
\[
\left[ c \left( \frac{\delta}{\| x^1 \|_u} \right)^{\frac{a}{2}} \right] \| v \|_{r-x}
\]
\[
= m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x} \leq (13) \text{ with } v = \frac{s - u}{a+s}
\]
\[
\leq m_{\hat{v}} \| r_\alpha (B^* B) \|_{r-x} \leq (13) \text{ with } v = \frac{s - u}{a+s}
\]
Thus, there exists \( C_2 \hat{=} m_{\hat{v}} c_\mu (c + 1)(M + 1) \) such that
\[
\| x_\alpha - x^1 \|_r \leq C_2 \| x^1 \|_u \delta^{\frac{a}{2}}
\]
Finally, from (33) and (34) it follows that there exists \( C \hat{=} C_1 + C_2 \) such that
\[
\| x_\alpha - x^1 \|_r \leq C \| x^1 \|_u \delta^{\frac{a}{2}}
\]
as we wanted to show.

Regarding the estimate (32) for the total error in the previous theorem, it is important to note the following: if \( r > 0 \) then the order of convergence that we obtain is worse than the one obtained in theorem 2.10 (see (23)), but now this order is obtained in the stronger \( \| \cdot \|_r \) norm. On the other hand if \( r \leq 0 \), then \( \| \cdot \|_r \) is weaker than \( \| \cdot \| \) and therefore (32) provides an estimate for the total error in a norm which is weaker than the norm in \( X \). However, in this case, it is important to note that the order \( O(\delta^{\frac{a}{2r}}) \) in (32) is now better than the one obtained in (23).

It is worth noting here that the parameter choice rule (31) requires the explicit knowledge of the degree of regularity \( a \) of \( x^1 \). However, the following result shows that convergence can also be obtained in the norm \( \| \cdot \|_r \) when the parameter choice rule is chosen in the form \( \alpha = c \delta^r \) taking any value within a certain interval.

**Theorem 3.3.** Let \( X, Y, T, T^1, L, (X_\alpha)_{\alpha \in R}, s \geq 0, a > 0, B = TL^{-s}, g_\alpha, r_\alpha, R_\alpha = L^{-s} g_\alpha (B^* B) B^*, y \in D(T^1), y^0 \in Y, \| y - y^0 \| \leq \delta, u \in [0, a + 2s], x^1 = T^1 y \in X_\alpha. \)
there exist constants $\sigma$, all as in theorem 2.10. Let $r \in [-a, \min\{u, s\}]$ and suppose that the parameter choice rule $\alpha = c \delta^r$ where $c \in (0, \frac{2(a+s)}{a+r}]$. Then,
\[
\|x_\alpha^\delta - x^\delta\|_r = O(\delta^r),
\]
where $\sigma = \min\{1 - \frac{2(a+r)}{2(a+s)}, \frac{2(a-r)}{2(a+s)}\}$. The optimal order of convergence is obtained when $\epsilon$ is chosen to be $\epsilon = \frac{2(a+s)}{a+r}$. In which case the order of convergence (32) of theorem 3.2 is obtained.

**Proof.** Following similar steps as in the proof on theorem 3.2, it immediately follows that there exist constants $C_1$ and $C_2$ such that
\[
\|x_\alpha^\delta - x_\alpha\|_r \leq C_1 \alpha^{-\frac{\|x^\delta\|_r}{\|x^\delta\|_r}} \text{ and } \|x_\alpha - x^\delta\|_r \leq C_2 \alpha^{-\frac{\|x^\delta\|_r}{\|x^\delta\|_r}}.
\]
Since $\alpha = c \delta^r$, it then follows that
\[
\|x_\alpha^\delta - x_\alpha\|_r \leq C_1 \delta^{1-\frac{\|x^\delta\|_r}{\|x^\delta\|_r}} \tag{35}
\]
and
\[
\|x_\alpha - x^\delta\|_r \leq C_2 \delta^{\frac{\|x^\delta\|_r}{\|x^\delta\|_r}}. \tag{36}
\]
From (35) and (36), it follows that
\[
\|x_\alpha^\delta - x^\delta\|_r = O(\delta^\sigma),
\]
where $\sigma = \min\{1 - \frac{2(a+r)}{2(a+s)}, \frac{2(a-r)}{2(a+s)}\}$. Also from (35) and (36) we have that the order of convergence is optimal when $\epsilon$ is chosen such that
\[
1 - \frac{\epsilon(a+r)}{2(a+s)} = \frac{\epsilon(u-r)}{2(a+s)},
\]
that is, for $\epsilon = \frac{2(a+s)}{a+r}$, in which case $\sigma = \frac{a+r}{a+s}$. \hfill \Box

It is important to note now that the results of theorems 3.1 and 3.2 are obtained for particular choices of the parameters in theorem 3.3. In fact, if $r = 0$, then theorem 3.3 yields the convergence result of theorem 3.1, while for $\epsilon = \frac{2(a+s)}{a+r}$ the convergence result of theorem 3.2 is obtained.

In the next theorem, we show that the optimal order of convergence in theorem 3.1 can also be achieved under the assumption of a source condition on $x^\delta$, associated with the restriction of the operator $T$ to the space $X_t$ of the Hilbert scale induced by $L$ for some $s \geq 0$.

**Theorem 3.4.** Let $\mathcal{X}$, $\mathcal{Y}$, $T$, $T^\dagger$, $L$, $(x_t)_{t \in \mathbb{R}}$, $s \geq 0$, $a > 0$, $\mu_0 \geq 1$, $B = TL^{-r}$, $g_\alpha$, $r_\alpha$, $R_\alpha = L^{-1}g_\alpha(B^tB)B^t$, $y \in \mathcal{D}(T^\dagger)$, $y^\delta \in \mathcal{X}$, $\|y - y^\delta\| \leq \delta$, $x^\dagger = T^\dagger y$, $x_\alpha = R_\alpha y$ and $x_\alpha^\delta = R_\alpha y^\delta$, all as in theorem 2.10. Suppose that $x^\dagger \in \mathcal{R}(L^{-1}T^*T_{y_\alpha})$ for some $u \in (s, 2\mu_0(a+s) - a)$ and that the regularization parameter $\alpha$ is chosen as
\[
\alpha = c \left( \frac{\delta}{\|x^\dagger\|_u} \right)^{-\frac{2(a+s)}{a+r}}, \tag{37}
\]
where $c > 0$. Then, there exists a constant $C$ (which depends on $a$ and $s$ but not on $u$) such that the following estimate for the total error holds:
\[
\|x_\alpha^\delta - x^\delta\| \leq C\delta^{\frac{a+r}{a+s}}. \tag{38}
\]
Proof. Consider the operator
\[ T_{ix} : (X_x, \| \cdot \|) \rightarrow Y. \] (39)
Observing that \( x \in X_x \), \( y \in Y \), we have
\[ \langle x, L^{-2s}T^*y \rangle_s = \langle L^*x, L^{-2s}T^*y \rangle \\
= \langle x, T^*y \rangle \\
= \langle Tx, y \rangle. \]
It then follows that the adjoint \( T^\# \) of the operator \( T_{ix} \) defined in (39) is given by \( T^\# = L^{-2s}T^* \). Hence, the source condition \( x^1 \in \mathcal{R}(L^{-2s}T^*T_{ix}) \) can also be written as \( x^1 = (T^2T)\tilde{m}v \) for some \( v \in X_x \).

On the other hand,
\[ R_u = L^{-s}g_u(B^*B)B^* \]
\[ = L^{-s}L^{-s}T^*T^*g_u(TL^{-s}L^{-s}T^*) \]
\[ = T^\#g_u(TT^\#) \]
\[ = g_u(T^\#T)T^\#, \] (40)
and therefore the family of operators \( R_u \) constitutes a spectral regularization for the operator \( T_{ix} \) given in (39).

Observe now that
\[ \| x^0_k - x^1 \| \leq \| x^0_k - x^1 \| \tilde{m} \| x^0_k - x^1 \| \tilde{m}^\# \]
\[ \leq m^{-1}\| T(x^0_k - x^1) \| \tilde{m} \| x^0_k - x^1 \| \tilde{m}^\#, \] (41)
where the first inequality follows from (7) with \( q = -a \) and \( r = 0 \) and the second one from (12).

For the first factor on the RHS of (41), we have the estimate
\[ \| T(x^0_k - x^1) \| \leq k\tilde{c} + \tilde{c} \alpha^{\tilde{a}+1/2}, \]
with \( \tilde{u} \approx \frac{\mu_ar}{\tilde{a}2}, \tilde{c} = \| v \| \) and \( k \) as in (24), where the last inequality follows immediately from (40) and from theorems 4.2 and 4.3 in [6] (note that \( 0 < \tilde{u} \leq \mu_a \frac{1}{2} \)). Then, with \( \alpha \) as in (37), it follows that
\[ \| T(x^0_k - x^1) \| \leq k\tilde{c} + \tilde{c} \| x^0_k \|^{\tilde{a}+1/2} (\delta^{\tilde{a}+1/2} \tilde{a}^{\tilde{a}+1/2} ) \]
\[ = (k + \tilde{c} \| x^0_k \|^{\tilde{a}+1/2}) \delta \]
\[ \leq \tilde{C}\delta, \] (43)
where \( \tilde{C} \approx k + \tilde{c}(1 + \gamma^{(1-2\tilde{a})\alpha})\| x^0_k \|^{\tilde{a}+1} \). Note here that \( \tilde{C} \) is independent of \( u \).

On the other hand, for the second factor in (41), from corollary 4.4 in [6] with \( \mu = \frac{a^{-1}}{\Sigma(a+1)} \), we get the estimate
\[ \| x^0_k - x^1 \| \leq c \delta \tilde{m} = c \delta \tilde{m}^\#, \] (45)
where \( c > 0 \).

Finally, with the estimates (42) and (45) in (41), we obtain
\[ \| x^0_k - x^1 \| \leq m^{-1}(\tilde{C}\delta \tilde{m} + c \delta \tilde{m}) = \tilde{C}\delta \tilde{m}, \]
where \( \tilde{C} \approx m^{-1}\tilde{C} \tilde{m} \). This concludes the proof. \( \square \)
In the next theorem, we will show that under the same conditions of theorem 3.4, with the additional hypotheses that the operators $L^{-1}$ and $T^*T$ commute, it is possible to obtain the same order of convergence as in (38), but now for a larger range of values of $u$.

**Theorem 3.5.** Let $X, Y, T, T^*, L, s \geq 0$, $u > 0$, $\mu_0 \geq 1$, $B = TL^{-1}$, $g_0$, $r_0$. $R_\alpha = L^{-\alpha}g_0(B^*B)B^*$, $y \in D(T^*)$, $y^\delta \in Y$, $\|y - y^\delta\| \leq \delta$, $x^\delta = T^*y$, $x_\alpha = R_\alpha y$, $x_\alpha^\delta = R_\alpha x^\delta$ and $\alpha = \alpha(\delta)$, all as in theorem 2.10. Suppose also that the operators $L^{-1}$ and $T^*T$ commute and that $x^\delta \in \mathcal{R}(B^*B)^\perp$ for some $u \in [0, 2\mu_0(a + s)]$. Then, there exists a constant $C$ (which depends on $a$ and $s$ but not on $u$) such that the following estimate for the total error holds:

$$
\|x_\alpha^\delta - x^\delta\| \leq C\delta^{\frac{\alpha}{m}}.
$$

**Proof.** To prove this result, we will follow similar steps as those in the previous theorems, proceeding to estimate the error due to noise and the regularization error separately. Just like in theorem 3.1, without loss of generality we will assume that $y \in \mathcal{R}(T)$. For the error due to noise, with the same proof as in theorem 3.1, from (25) we have that

$$
\|x_\alpha^\delta - x_\alpha\| \leq C_1\delta^{\frac{\alpha}{m}}
$$

(46)

where $C_1 = k\,m^{\frac{\alpha}{m}}$, with $k$ as in (24) and $m$ as in (12).

On the other hand, since $L^{-1}$ commutes with $T^*T$, it follows that $L^{-\alpha}$ commutes with $B^*B$ and therefore, with any function of $B^*B$. Let $v \in X$ such that $x^\delta = (B^*B)^\perp v$. Then, for the regularization error we have that

$$
\|x_\alpha - x^\delta\| = \|R_\alpha y - x^\delta\|
= \|L^{-\alpha}g_0(B^*B)B^*y - x^\delta\|\hspace{1cm} (\text{since } B^*y = B^*BL^x x^\delta)
= \|L^{-\alpha}g_0(B^*B)B^*BL^x I \cdot x^\delta\|\hspace{1cm} (\text{since } L^{-\delta} \text{ commutes with } T^*T)
= \|g_0(B^*B)B^*BL^x I \cdot (B^*B)^\perp v\|\hspace{1cm} (\text{since } x^\delta \in \mathcal{R}(B^*B)^\perp)
= \|r_0(B^*B)(B^*B)^\perp v\|
= \|(B^*B)^\perp r_0(B^*B)v\|
\leq c_2\alpha^{\frac{\alpha}{m}}\|v\|\hspace{1cm} \text{(by (20) with } \bar{\mu} = \frac{\mu}{2(a + s)})
\leq C_2\alpha^{\frac{\alpha}{m}}\|v\|.
$$

(47)

Thus,

$$
\|x_\alpha - x^\delta\| \leq C_2\alpha^{\frac{\alpha}{m}}.
$$

(48)

Finally, from (46) and (47), it follows that

$$
\|x_\alpha^\delta - x^\delta\| \leq C_1\delta^{\frac{\alpha}{m}} + C_2\alpha^{\frac{\alpha}{m}}
= C_1\delta^{\frac{\alpha}{m}}\|x^\delta\|^{\frac{\alpha}{\mu}} + C_2\delta^{\frac{\alpha}{m}}\|x^\delta\|^{\frac{\alpha}{\mu}}
\leq C\delta^{\frac{\alpha}{m}}.
$$

\(\square\)

In the table below and in figure 2 we illustrate the restrictions on the parameter $u$ and the source condition for $x^\delta$ guaranteeing the order of convergence given in (23). These results were obtained in theorems 2.10, 3.4 and 3.5, respectively.
Let $x \in X_s$ where the last equality follows immediately from lemma 2.6.

\[
\begin{array}{|c|c|}
\hline
\text{Source condition} & \text{Restriction on } u \\
\hline
x^I \in \mathcal{X}_s & 0 \leq u \leq a + 2s \\
x^I \in \mathcal{R}((L^{-2s}T^s L_x \big|_{\mathcal{X}_s}) \bigg|_{\mathcal{X}_s}) & s < u \leq 2\mu_0(a+s) - a \\
x^I \in \mathcal{R}((B^* B)^{\frac{1}{2}}) & 0 \leq u \leq 2\mu_0(a+s) \\
\hline
\end{array}
\]

In the following proposition, a relation between the source sets of theorems 2.10 and 3.4 is shown.

**Proposition 3.6.** Let $\mathcal{X}, \mathcal{Y}, T, L, (\mathcal{X}_s)_{s \in \mathbb{R}}$, $s \geq 0$, $a > 0$ and $B = TL^{-1}$, all as in theorem 3.4. Then, for every $u \in [s, a + 2s]$ there holds

\[
\mathcal{X}_u \subseteq \mathcal{R}((L^{-2s}T^s T_x \big|_{\mathcal{X}_s}) \bigg|_{\mathcal{X}_s}).
\]

(49)

For $u = a + 2s$ the inclusion in (49) is in fact an equality.

**Proof.** Let $T^s = L^{-2s}T^*$ be the adjoint of the operator $T_{|X_s}$ as defined in (39). Then, for every $x \in \mathcal{X}_s$ we have that

\[
\| (T^s T_{|X_s})^{1/2} x \|_s^2 = \langle T^s T_{|X_s} x, x \rangle_s = \| T_{|X_s} x \|^2_s.
\]

From this equality and (12), it follows that

\[
m \| x \|_{u-a} \leq \| (T^s T_{|X_s})^{1/2} x \|_s \leq M \| x \|_{u-a} \quad \forall x \in \mathcal{X}_s.
\]

(50)

On the other hand, note that

\[
\mathcal{D}((T^s T_{|X_s})^{-1/2}) = \mathcal{R}((T^s T_{|X_s})^{1/2}) = \mathcal{R}(T^s) = \mathcal{R}(L^{-2s}T^*) = \mathcal{X}_{u+2s},
\]

(51)

where the last equality follows immediately from lemma 2.6.

Now, using (50), (51) and a standard duality argument, it follows easily that

\[
\frac{1}{M} \| x \|_{u+2s} \leq \| (T^s T_{|X_s})^{-1/2} x \|_s \leq \frac{1}{m} \| x \|_{u+2s} \quad \forall x \in \mathcal{X}_{u+2s}.
\]

(52)

From (51) and (52), the use of the Heinz inequality (theorem 2.4) for the operators $L^{v+2s}$ and $(T^s T_{|X_s})^{-1/2}$ allows us to conclude that for every $v \in [0, 1]$ there holds

\[
\mathcal{D}(L^{v(u+2s)}) = \mathcal{D}((T^s T_{|X_s})^{-v/2})
\]

(53)

and

\[
M^{-v} \| L^{v(u+2s)} x \| \leq \| (T^s T_{|X_s})^{-v/2} x \| \leq m^{-v} \| L^v (a+2s) x \| \quad \forall x \in \mathcal{D}(L^{v(u+2s)}).
\]
Finally we have that
\[ \mathcal{X}_u = \mathcal{D}(L^u) \]
\[ = \mathcal{D}(L^{(u+2s)}) \]
\[ = \mathcal{D}(\{(T^2T_{\mathcal{X}_u})^{-\frac{s}{1+s}}\}) \]
\[ = \mathcal{R}(\{(T^2T_{\mathcal{X}_u})^{-\frac{s}{1+s}}\}) \subset \mathcal{R}(\{(L^{-2s}T^*T_{\mathcal{X}_u})^{-\frac{s}{1+s}}\}), \]
which proves the first part of the proposition.

For the second part, note that if \( u = a + 2s \), then
\[ \mathcal{X}_u = \mathcal{X}_{a+2s} \]
\[ = \mathcal{D}(\{(T^2T_{\mathcal{X}_u})^{-1/2}\}) \]
\[ = \mathcal{R}(\{(T^2T_{\mathcal{X}_u})^{1/2}\}) \]
\[ = \mathcal{R}(\{(L^{-\frac{1}{2}}T^*T_{\mathcal{X}_u})^{\frac{1}{2+s}}\}) \]
This completes the proof of the proposition. \( \square \)

It is worth noting that the inclusion in (49) reveals that the source condition \( x^i \in \mathcal{R}(\{(L^{-2s}T^*T_{\mathcal{X}_u})^{-\frac{s}{1+s}}\}) \) in theorem 3.4 is less restrictive than the source condition \( x^i \in \mathcal{X}_u \) of theorem 2.10 for values of \( u \in [s, a+2s] \). Therefore, the latter theorem can now be seen as a corollary of theorem 3.4. Moreover, note that since \( \mu_a \geq 1 \), theorem 3.4 is valid for \( u \) in a set which is larger than the one for which theorem 2.10 holds. In light of this observation, it is then reasonable to question the relevance of theorem 2.10. An answer to this question is immediately obtained by observing that the source condition \( x^i \in \mathcal{X}_u \), although more restrictive than the condition \( x^i \in \mathcal{R}(\{(L^{-2s}T^*T_{\mathcal{X}_u})^{-\frac{s}{1+s}}\}) \) is, in general, easier to verify since it involves only the operator \( L \) while the latter involves both \( L \) and \( T \). On the other hand, if the operators \( L^{-1} \) and \( T^*T \) commute, then there exist close connections between the source conditions in theorems 2.10, 3.4 and 3.5. We will establish these connections in corollary 3.11. An extension of the second part of theorem 2.8 namely identity (15) for values of \( v > 1 \), will be previously needed. We will obtain such extension in theorem 3.10. A few previous results, which are presented in the next three lemmas, will be needed.

**Lemma 3.7.** Let \( \mathcal{X}, \mathcal{Y}, T, L, (\mathcal{X}_s) \in \mathbb{R} \) and \( s \geq 0 \), all as in proposition 3.6. Suppose also that there exist positive constants \( m, M \) with \( 0 < m \leq M < \infty \) and \( a \in \mathbb{R}^+ \) such that (12) holds, i.e.
\[ m\|x\|_{-a} \leq \|Tx\| \leq M\|x\|_{-a} \quad \forall \ x \in \mathcal{X}; \] (54)
then:

(i) \( \mathcal{R}(T^*T) \subset \mathcal{X}_{2a}; \)

(ii) \( \mathcal{R}(L^{-2s}T^*T) \subset \mathcal{X}_{2(a+1)} \).

If \( L^{-1} \) and \( T^*T \) commute, then equality holds in both inclusions above.
Proof. To prove (i), observe that since \( M \subset X \) \( \forall t \) and \( \overline{M} = X \), it follows immediately that \( \overline{M} = X \). Suppose now that \( x \in R(T^*T) \). Then, from lemma 2.6 \( x \in X \). Hence, there exists a sequence \( \{x_n\} \subset X \) such that \( \|x_n - x\| \to 0 \) and therefore also \( \|x_n - x\| \to 0 \). Then, \( \|TL^{2a}x_n\| \leq M \|L^{2a}x_n\| \leq M \|L^{2a}x_n\| < \infty \). Thus, the sequence \( \{TL^{2a}x_n\} \) is bounded in \( Y \) and therefore there exist \( y \in Y \) and a subsequence of \( \{x_n\} \) (also denoted by \( \{x_n\} \)) such that \( TL^{2a}x_n \to y \). Finally, since the operator \( TL^{2a} \) is closed, we have that \( x \in D(L^{2a}) = X \) and, moreover, \( TL^{2a}x = y \). Thus, \( R(T^*T) \subset X \), which proves (i).

Suppose now that \( L^{-1} \) and \( T^*T \) commute and let \( x \in X \). Then, \( L^{a}x \in X \) and therefore there exists \( x_1 \in X \) such that \( L^{a}x = (T^*T)^{1/2}x_1 \). Then, \( x = L^{-a}(T^*T)^{1/2}x_1 = (T^*T)^{1/2}L^{-a}x_1 \), where the last equality holds by virtue of the commutativity of \( L^{-1} \) and \( T^*T \). Now, since \( L^{-a}x_1 \in X \) = \( R((T^*T)^{1/2}) \), it follows that there exists \( w \in X \) such that \( L^{-a}x_1 = (T^*T)^{1/2}w \). Finally then, \( x = T^*Tw \in R(T^*T) \) and hence equality holds in (i).

To prove (ii), let \( x \in R(L^{-2T^*T}) \). Then there exists \( x_0 \in X \) such that \( L^{-2T^*T}x_0 = x \). But from (i) it follows that \( T^*Tx_0 \in X \) and therefore \( L^{-2T^*T}x_0 \in X \). On the other hand if \( L^{-1} \) and \( T^*T \) commute and \( x \in X \), then there exists \( L^2x \in X \) and \( L^2x \in X \). Since in this case equality in (i) holds, it then follows that \( L^2x \in R(T^*T) \). Hence, there exists \( x_0 \in X \) such that \( L^2x = T^*Tx_0 \), and therefore \( x = L^{-2T^*T}x_0 \in R(L^{-2T^*T}) \). This concludes the proof of the lemma.

Lemma 3.8. Let \( X, Y, T, L, (X_i)_{i \in \mathbb{R}}, s \geq 0, a > 0, m, M, all as in lemma 3.7 \), and \( B = TL^{-s} \) as in theorem 3.1. If \( L^{-1} \) and \( T^*T \) commute then:

(i) \( R(B^*B) = X_{2(a+s)} \);  
(ii) \( m^2 \|x\|_{-2(a+s)} \leq \|B^*Bx\| \leq M^2 \|x\|_{-2(a+s)} \ \forall \ x \in X \);  
(iii) \( M^{-2} \|x\|_{2(a+s)} \leq \|B^*B^{-1}x\| \leq m^{-2} \|x\|_{2(a+s)} \ \forall \ x \in X_{2(a+s)} \).

Proof. Note that (i) follows immediately from the previous lemma. To prove (ii) observe that for every \( x \in X \) we have

\[
B^*Bx = L^{-2T^*T}x = T^*TL^{-2a}x = (T^*T)^{1/2}L^{-2a}L^{-2(a+s)}x.
\]

Thus,

\[
\|B^*Bx\| = \|(T^*T)^{1/2}L^{-2a}L^{-2(a+s)}x\| \leq M\|(T^*T)^{1/2}L^{-2(a+s)}x\| \leq M\|Bx\| \leq m \|x\|_{2(a+s)} \quad \text{(from (12))}
\]

Similarly, by using the inequality \( m \|x\|_{-a} \leq \|Tx\| \), it follows that \( B^*Bx \) \( \|x\|_{-2(a+s)} \leq B^*Bx \). This completes the proof of (ii).
Inverse Problems 28 (2012) 055005
G L Mazzi fieri and R D Spies

To prove (iii), let \( x \in \mathcal{X}_{2(a+s)} \). Then,

\[
\| (B^*B)^{-1}x \| = \sup_{x \in \mathcal{X}_{2(a+s)}} |\langle (B^*B)^{-1}x, \bar{x} \rangle | \\
= \sup_{x \in \mathcal{X}_{2(a+s)}} |\langle x, (B^*B)^{-1}\bar{x} \rangle | \\
= \sup_{x \in \mathcal{X}_{2(a+s)}} |\langle x, z \rangle | \\
\leq \sup_{x \in \mathcal{X}} |\langle x, z \rangle | \quad \text{(from (ii))}
\]

A similar reasoning allows us to prove that \( M^{-2} \| x \|_{2(a+s)} \leq \| (B^*B)^{-1}x \| \). This concludes the proof of the lemma.

In the following lemma, it is proved that, under the hypothesis of commutativity of the operators \( A \) and \( L \), the Heinz inequality (theorem 2.4) is also valid for \( v > 1 \).

**Lemma 3.9.** Let \( A \) and \( L \) be two unbounded, self-adjoint, strictly positive operators on a Hilbert space \( \mathcal{X} \). Suppose also that \( \mathcal{D}(A) \subset \mathcal{D}(L) \), \( A \) and \( L \) commute on \( \mathcal{D}(A) \) and \( \| Lx \| \leq \| Ax \| \) for every \( x \in \mathcal{D}(A) \). Then, for every \( k \geq 0 \) it follows that \( \mathcal{D}(A^k) \subset \mathcal{D}(L^k) \) and \( \| L^k x \| \leq \| A^k x \| \) for every \( x \in \mathcal{D}(A^k) \).

**Proof.** If \( 0 \leq k \leq 1 \), the result is true by virtue of the Heinz inequality (theorem 2.4). Suppose then that \( k > 1 \). We will prove first that the result is true for all \( k \in \mathbb{N} \), that is, we will first show, by induction, that \( \mathcal{D}(A^k) \subset \mathcal{D}(L^k) \) and \( \| L^k x \| \leq \| A^k x \| \) \( \forall x \in \mathcal{D}(A^k) \), \( \forall n \in \mathbb{N} \). For that, let \( n = 2 \) and \( x \in \mathcal{D}(A^2) \). Since \( x \in \mathcal{D}(A^2) \subset \mathcal{D}(A) \subset \mathcal{D}(L) \), there exists \( w = Lx \). On the other hand, \( x \in \mathcal{D}(A^2) \). Thus, \( z = L^2 x = A^2 x = L^2 x \).

Then, \( w \in \mathcal{D}(A) \subset \mathcal{D}(L) \) and therefore there exists \( r \in \mathcal{X} \) such that \( r = Lw = L^2 x \). Hence, \( x \in \mathcal{D}(L^2) \). We also proved that \( \mathcal{D}(A^2) \subset \mathcal{D}(L^2) \). Also, for \( x \in \mathcal{D}(A^2) \) we have that \( \| L^2 x \| \leq \| A^2 x \| \) \( \leq \| L^2 x \| \). Since \( \mathcal{D}(A^2) \subset \mathcal{D}(L^2) \), there exists \( w = L^2 x \). On the other hand, \( x \in \mathcal{D}(A^2) \) and by the inductive hypothesis \( \mathcal{D}(A^2) \subset \mathcal{D}(L^2) \). Then, there exists \( z = L^2 x \). Thus,

\[
 z = L^2 x = A^2 x = L^2 x.
\]

Then, \( w \in \mathcal{D}(L) \) and therefore there exists \( r = Lw = LL^2 x = L^{n+1} x \). Hence, \( x \in \mathcal{D}(L^{n+1}) \). Also if \( x \in \mathcal{D}(A^{n+1}) \), then \( \| L^{n+1} x \| = \| L^k L^k x \| \leq \| A^k L^k x \| = \| L^k A^k x \| \leq \| A^{n+1} x \| \).

We have then proved that for every \( n \in \mathbb{N} \)

\[
 \mathcal{D}(A^k) \subset \mathcal{D}(L^k) \quad \text{and} \quad \| L^k x \| \leq \| A^k x \| \quad \forall x \in \mathcal{D}(A^k) \quad \text{for} \quad k \geq 0.
\]

(55)
Suppose now that \( k \in \mathbb{R}^+ \setminus \mathbb{N} \) and define \( n = \lceil k \rceil \) (where \( \lceil \cdot \rceil \) denotes the ‘ceiling’ function). Since \( n \in \mathbb{N} \), from (55) we have that \( D(A^*) \subset D(L^\nu) \) and \( \|L^\nu x\| \leq \|A^\nu x\| \). Now, by using theorem 2.4 with \( L \) and \( A \) replaced by \( L^\nu \) and \( A^\nu \) and \( \nu = \frac{k}{|k|} \), it follows that \( D(A^*) \subset D(L^\nu) \) and \( \|L^\nu x\| \leq \|A^\nu x\| \) \( \forall x \in D(A^*) \), that is,

\[
D(A^*) \subset D(L^\nu) \quad \text{and} \quad \|L^\nu x\| \leq \|A^\nu x\| \quad \forall x \in D(A^*).
\]

\( \square \)

Having proved the three previous lemmas, we are now ready to prove an extension of the identity (15) of theorem 2.8 which will allow us to show the relationships between the source conditions of theorems 2.10 and 3.5, that is, conditions of the form \( x^\dagger \in \mathcal{X}_0 \) and \( x^\dagger \in \mathcal{R}((B^*B)^{\frac{n}{2}}) \) for the case in which \( L^{-1} \) and \( T^*T \) commute.

**Theorem 3.10.** Let \( T : \mathcal{X} \rightarrow \mathcal{Y} \) be a linear continuous operator between the Hilbert spaces \( \mathcal{X} \) and \( \mathcal{Y} \), \( L \) a linear, densely defined, unbounded and strictly positive operator on \( \mathcal{X} \), and \( (\mathcal{X}_i)_{i \in \mathbb{R}} \) the Hilbert scale induced by \( L \) over \( \mathcal{X} \). Let also \( s \) be a positive constant, \( B = TL^{-s} \) and suppose that there exist positive constants \( a, m \) and \( M \) such that (12) holds. Assume also that \( L^{-1} \) and \( T^*T \) commute. Then, for every \( r > 0 \) we have that

\[
\mathcal{R}((B^*B)^{\frac{n}{2}}) = \mathcal{X}_{2s+a+2s}.
\]

**Proof.** First note that from lemma 3.8 (i) it follows that \( D((B^*B)^{-1}) = \mathcal{X}_{(a+s)} = D(L^{2(a+s)}) \). On the other hand, since the operators \( L^{-1} \) and \( T^*T \) commute, then \( T^*T \) and \( L^{-r} \) also commute for every \( r > 0 \) (see [7], p. 140). Then, the operators \( B^*B = L^{-s}T^*TL^{-s} \) and \( L^{-2(a+s)} \) commute and therefore their respective inverses also commute. From lemma 3.8 (iii) and lemma 3.9, it then follows that

\[
D(((B^*B)^{-1})^{\nu}) = D((L^{2(a+s)})^{\nu}) \quad \forall \nu \geq 0,
\]

that is,

\[
\mathcal{R}((B^*B)^{\nu}) = \mathcal{X}_{2s+a+2s}.
\]

\( \square \)

The following corollary shows the relation between the source conditions of theorems 2.10, 3.4 and 3.5.

**Corollary 3.11.** Let \( \mathcal{X}, \mathcal{Y}, T, L, (\mathcal{X}_i)_{i \in \mathbb{R}}, s, a \) and \( B \), all as in theorem 3.10. Then,

(i) \( \mathcal{X}_u = \mathcal{R}((B^*B)^{\frac{n}{2}}) \quad \forall u \geq 0 \).

(ii) \( \mathcal{R}((L^{-2s}T^*T)^{\frac{n}{2}}) \subset \mathcal{R}((L^{-2s}T^*T_{\mathcal{X}})^{\frac{n}{2}}) \quad \forall u \in [s, a + 2s] \).

**Proof.** Part (i) follows immediately from theorem 3.10 with \( \nu = \frac{u}{s+a} \). To prove (ii) note that if \( u \in [s, a + 2s] \), then

\[
\mathcal{R}((L^{-2s}T^*T)^{\frac{n}{2}}) = \mathcal{R}((B^*B)^{\frac{n}{2}}) \quad \text{(since } L^{-s}T^*T \text{ commute)}
\]

\[
= \mathcal{X}_u \quad \text{(by (i))}
\]

\[
\subset \mathcal{R}((L^{-2s}T^*T_{\mathcal{X}})^{\frac{n}{2}}) \quad \text{(by proposition 3.6)}.
\]

Hence,

\[
\mathcal{R}((L^{-2s}T^*T)^{\frac{n}{2}}) \subset \mathcal{R}((L^{-2s}T^*T_{\mathcal{X}})^{\frac{n}{2}}),
\]

as we wanted to prove. \( \square \)
Remark 3.12. Under the hypothesis that the operators $L^{-1}$ and $T^*T$ commute, corollary 3.11 implies that for $u \in [s, a + 2s]$ the source condition $x^T \in \mathcal{R}(B^*B)^{\infty\infty}$ of theorem 3.5 is more restrictive than the source condition $x^T \in \mathcal{R}(L^{-2}T^*T_y)^{\infty\infty}$ of theorem 3.4. However, it is important to point out here that theorem 3.5 is valid for a set of values of $u$ which is larger than the one for which theorem 3.4 is valid. In particular, theorem 3.5 is valid for values of $u \in (2\mu_a(a + s) - a, 2\mu_a(a + s)]$ (for which theorem 3.4 is not valid), thus allowing us to obtain better orders of convergence.

4. Main results

4.1. Multiple Hilbert scales

In this section, we will first introduce the concept of a multiple (or vectorial) Hilbert scale. Then, we will define a regularization method in these multiple Hilbert scales and prove several convergence theorems, some of which generalize results obtained in the previous section.

Let $T$ be a linear continuous operator between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. Consider also $N$ linear, densely defined, unbounded, strictly positive, self-adjoint operators, with open dense domains:

$$L_i : \mathcal{D}(L_i) \subset \mathcal{X} \rightarrow \mathcal{X}, \quad i = 1, 2, ..., N.$$  \hspace{1cm} (57)

Thus, each $L_i$ is a closed operator on $\mathcal{X}$ satisfying: $\mathcal{D}(L_i) = \mathcal{D}(L_i^*)$ is dense in $\mathcal{X}$, and $L_i$ is dense in $\mathcal{X}$, $(L_i x, y) = (x, L_i^* y)$ for every $x, y \in \mathcal{D}(L_i)$ and there exists a positive constant $\gamma_i$ such that $\langle L_i x, x \rangle \geq \gamma_i \|x\|^2$ for every $x \in \mathcal{D}(L_i)$.

In what follows, we will obtain regularized solutions of the ill-posed problem $Tx = y$ by means of the simultaneous use of the $N$ Hilbert scales induced on $\mathcal{X}$ by the operators $L_i$, $1 \leq i \leq N$. The motivation for this development comes mainly from the idea of combining the advantages of the use of general penalizing terms in Tikhonov–Phillips-type methods (see [16]) with the adaptive virtues that regularization in Hilbert scales possesses in regard to the order of convergence of the total error as a function of the degree of regularity of the exact solution. In order to do that we will previously need to extend some of the concepts and definitions that were introduced in section 2.

For each index $i$, $1 \leq i \leq N$, consider the set $\mathcal{M}_i$ of all the elements $x \in \mathcal{X}$ for which all natural powers of $L_i$ are defined, i.e. $\mathcal{M}_i \equiv \bigcap_{s=1}^{\infty} \mathcal{D}(L_i^s)$. As seen in section 2, $\mathcal{M}_i$ is dense in $\mathcal{X}$, the powers $L_i^s$ are well defined on $\mathcal{M}_i$ for every $s \in \mathbb{R}$ and $\mathcal{M}_i = \bigcap_{s \in \mathbb{R}} \mathcal{D}(L_i^s)$. Now, for each $i = 1, 2, ..., N$ and for each $s \in \mathbb{R}$, we define the mapping $\langle \cdot, \cdot \rangle_{L_i^s} : \mathcal{M}_i \times \mathcal{M}_i \rightarrow \mathbb{C}$ as $\langle x, y \rangle_{L_i^s} = (L_i^s x, L_i^s y)$, $x, y \in \mathcal{M}_i$. Again, as seen in section 2, $(\cdot, \cdot)_{L_i^s}$ defines an inner product in $\mathcal{M}_i$, which induces the norm $\|x\|_{L_i^s} = \|L_i^s x\|$, and $L_i$ induces on $\mathcal{X}$ a Hilbert scale that we will denote with $(\mathcal{X}^i_{L_i^s})_{s \in \mathbb{R}}$. Here $\mathcal{X}^i_{L_i^s}$ is the completion of $\mathcal{M}_i$ in the $\|\cdot\|_{L_i^s}$-norm.

Let us now consider the Hilbert space $\mathcal{X}_N$ consisting of $N$ copies of $\mathcal{X}$, i.e. $\mathcal{X}_N \equiv \bigotimes_{i=1}^{N} \mathcal{X}$ with the usual inner product in a product space. With the operators $L_i$, $L_2$, ..., $L_N$ given in (57) we define the operator $\vec{L} : \mathcal{X}_N \rightarrow \mathcal{X}_N$ as

$$\mathcal{D}(\vec{L}) = \bigotimes_{i=1}^{N} \mathcal{D}(L_i), \quad \vec{L} = \text{diag}(L_1, L_2, \ldots, L_N),$$  \hspace{1cm} (58)

so that for $\vec{x} = (x_1, x_2, \ldots, x_N)^T \in \mathcal{D}(\vec{L})$ one has $\vec{L} \vec{x} = (L_1 x_1, L_2 x_2, \ldots, L_N x_N)^T$. Given the operator $\vec{L}$ defined as in (58) and $\vec{x} = (s_1, s_2, \ldots, s_N)^T \in \mathbb{R}^N$, the operator $\vec{L}$ is defined as $\vec{L} = \text{diag}(L_1^s, L_2^s, \ldots, L_N^s)$, with $\mathcal{D}(\vec{L}^s) = \bigotimes_{i=1}^{N} \mathcal{D}(L_i^s)$, so that for $\vec{x} = (x_1, x_2, \ldots, x_N)^T \in \mathcal{D}(\vec{L}^s)$

$$\vec{L} \vec{x} = (L_1^s x_1, L_2^s x_2, \ldots, L_N^s x_N)^T,$$  \hspace{1cm} (59)
Now, for every \( s = (s_1, s_2, \ldots, s_N)^T \in \mathbb{R}^N \) and \( x, y \in \mathcal{M} = \bigotimes_{i=1}^{N} \mathcal{M}_i \), we define the inner product \( \langle x, y \rangle_{\mathcal{M}} = \sum_{i=1}^{N} \langle x_i, y_i \rangle_{\mathcal{M}_i} \). It can be immediately seen that \( \langle \cdot, \cdot \rangle_{\mathcal{M}} \) defines an inner product in \( \mathcal{M} \), which in turn induces the norm \( \| x \|_{\mathcal{M}}^2 = \sum_{i=1}^{N} \| x_i \|_{\mathcal{M}_i}^2 \).

**Definition 4.1 (Multiple or vectorial Hilbert scale).** Let \( \hat{L} \) be as in (58), \( \bar{s} = (s_1, s_2, \ldots, s_N)^T \in \mathbb{R}^N \) and \( \mathcal{M} = \bigotimes_{i=1}^{N} \mathcal{M}_i \). The Hilbert space \( X_{\bar{s}}^{\mathcal{M}} \) is defined as the completion of \( \mathcal{M} \) with respect to the \( \| \cdot \|_{\mathcal{M}} \)-norm. The family of spaces \( \{ X_{\bar{s}}^{\mathcal{M}} \}_{\bar{s} \in \mathbb{R}^N} \) is called the vectorial Hilbert scale induced by \( \hat{L} \) over \( X^{\mathcal{M}} \). The operator \( \hat{L} \) is called a ‘generator’ of the Hilbert scale \( \{ X_{\bar{s}}^{\mathcal{M}} \}_{\bar{s} \in \mathbb{R}^N} \).

**Remark 4.2.** Since \( \hat{L} \) is diagonal, it can be easily seen that \( \Pi_{\bar{s}} X_{\bar{s}}^{\mathcal{M}} = X^{\mathcal{M}}_{\bar{s}} \) where \( \Pi_{\bar{s}} \) denotes the \( \bar{s} \)-th canonical projection of \( X^{\mathcal{M}} \) onto \( X \). Moreover, for any \( i \in \{ 1, 2, \ldots, N \} \), each one of them with the usual inherited inner product. Given \( \bar{s} \in \mathbb{R}^N \) we define the operator \( B : X^{\mathcal{M}} \rightarrow Y^{\mathcal{M}} \) as \( B = \hat{T} \hat{L}^{-\bar{s}} \), where \( \hat{T} : X^{\mathcal{M}} \rightarrow Y^{\mathcal{M}} \) is defined by \( \hat{T} = \text{diag}(T, T, \ldots, T) \). Thus, given \( \bar{x} \in X^{\mathcal{M}} \)

\[
\bar{B} \bar{x} = \hat{T} \hat{L}^{-\bar{s}} \bar{x} = (T L_{n_1}^{-s_1} x_1, T L_{n_2}^{-s_2} x_2, \ldots, T L_{n_N}^{-s_N} x_N)^T.
\]

From the properties of the operators \( L_i \), it follows immediately that the adjoint of \( \bar{B} \) is given by \( \bar{B}^* = \hat{T}^* \hat{L}^{-s} \), where \( \hat{T}^* = \text{diag}(T^*, T^*, \ldots, T^*) \). Thus, for every \( \bar{y} \in Y^{\mathcal{M}} \) we have that

\[
\bar{B}^* \bar{y} = \hat{L}^{-s} \hat{T}^* \bar{y} = (L_{n_1}^{-s_1} T^* y_1, L_{n_2}^{-s_2} T^* y_2, \ldots, L_{n_N}^{-s_N} T^* y_N)^T,
\]

and therefore for every \( \bar{x} \in X^{\mathcal{M}} \) there holds

\[
\bar{B}^* \bar{B} \bar{x} = \hat{L}^{-s} \hat{T}^* \hat{L}^{-s} \bar{x} = (L_{n_1}^{-s_1} T^* T L_{n_1}^{-s_1} x_1, L_{n_2}^{-s_2} T^* T L_{n_2}^{-s_2} x_2, \ldots, L_{n_N}^{-s_N} T^* T L_{n_N}^{-s_N} x_N)^T
\]

where \( B_i = T L_{n_i}^{-s_i} \) and \( B_i^* = L_{n_i}^{-s_i} T^* \) is the adjoint of the operator \( B_i \) (compare with the definition of \( B \) given in theorem 2.8). Note that the operators \( \bar{B}^* \bar{B} \) and \( B_i^* B_i \), for each index \( i, 1 \leq i \leq N \), are linear self-adjoint operators on the Hilbert spaces \( X^{\mathcal{M}} \) and \( X \), respectively. As such, for each one of them there exists a unique spectral family which allows them to be represented in terms of the integral of the identity with respect to the ‘operator-valued measure’ induced by that spectral family (see [2, 6, 7]). We will denote with \( \{ E_{\lambda}^{X_{\bar{s}}} \}_{\lambda \in \mathbb{R}} \) and \( \{ E_{\lambda}^{X_{\bar{s}}} \}_{\lambda \in \mathbb{R}} \) the spectral families of the operators \( \bar{B}^* \bar{B} \) and \( B_i^* B_i \), respectively (note that these families are partitions of the identity on the spaces \( X^{\mathcal{M}} \) and \( X \), respectively).

Let \( g : \mathbb{R} \rightarrow \mathbb{R} \) be a piecewise continuous function and consider the operators \( g(\bar{B}^* \bar{B}) : X^{\mathcal{M}} \rightarrow X^{\mathcal{M}} \) and \( g(B_i^* B_i) : X \rightarrow X \), \( 1 \leq i \leq N \). From (62), it can be easily proved that

\[
(g(\bar{B}^* \bar{B}) \bar{x})_i = g(B_i^* B_i)x_i,
\]

where \( \bar{x} = (x_1, x_2, \ldots, x_N)^T \in X^{\mathcal{M}} \).

The next theorem states a convergence result which generalizes theorem 3.1 to the case of multiple Hilbert scales.
Theorem 4.3. Let $T \in \mathcal{L}(X, Y)$ with $X$ and $Y$ Hilbert spaces, $L_i : D(L_i) \subset X \rightarrow X$, $1 \leq i \leq N$, linear densely defined, self-adjoint and strictly positive operators on $X$, each one of them with an open domain, $L_i \geq y_i$ for a constant $y_i > 0$, and let $\tilde{L} : X^N \rightarrow X^N$ be as defined in (58). Suppose also that for each index $i$, $1 \leq i \leq N$, there exist constants $m_i, M_i$, with $0 < m_i \leq M_i < \infty$, and $a_i > 0$, such that for every $x \in X$ the following condition holds:

$$m_i \|x\|_{x_i - \bar{\gamma}} \leq \|Tx\| \leq M_i \|x\|_{x_i - \bar{\gamma}}.$$  

(64)

Let $\bar{s} = (s_1, s_2, \ldots, s_N)^T \in \mathbb{R}_+^N$, $\bar{T} = \text{diag} (T, T, \ldots, T)$, $\bar{B} \doteq \bar{T} \bar{L}^{-1}$, $\bar{\gamma} = (\eta_1, \eta_2, \ldots, \eta_N)^T \in \mathbb{R}_+^N$ such that $\sum_{i=1}^N \eta_i = 1$. Also let $g_\alpha : [0, \|\bar{B}\|_\alpha^2] \rightarrow \mathbb{R}$, $\alpha > 0$, be a family piecewise continuous real-valued functions verifying the following conditions.

C1: For every $\lambda \in (0, \|\bar{B}\|_\alpha^2]$ there holds $\lim_{\lambda \rightarrow 0^+} g_\alpha(\lambda) = \frac{1}{\gamma^{1}}$.

C2: There exists a constant $\hat{c}$ such that $\forall \lambda \in (0, \|\bar{B}\|_\alpha^2]$ and $\forall \alpha > 0$ there holds $|g_\alpha(\lambda)| \leq \hat{c} \alpha^{-1}$.

C3: There exists $\mu_0 \geq 1$ such that if $\mu \in [0, \mu_0]$ then $\lambda^{\alpha} |r_\alpha(\lambda)| \leq c_\mu \alpha^{\mu}$ $\forall \lambda \in (0, \|\bar{B}\|_\alpha^2]$, where $c_\mu$ is a positive constant and $r_\alpha(\lambda) \doteq 1 - \lambda g_\alpha(\lambda)$.

For $y \in D(T^\alpha)$, $\delta > 0$ and $\gamma^\alpha \in Y$ with $\|y - \gamma^\alpha\| \leq \delta$, we define the regularized solution of the problem $Tx = y$ with data $\gamma^\alpha$, as

$$x^\alpha_{\delta} \doteq \bar{\gamma}^{\alpha} g_\alpha(\bar{B}_\alpha^T \bar{B}_\alpha)^{\gamma^\alpha},$$  

(65)

where $\gamma^\alpha$ is the observation vector defined as $\gamma^\alpha = (\gamma^1, \gamma^2, \ldots, \gamma^N)^T \in Y^N$. Suppose that for each index $i$, $1 \leq i \leq N$, there exists $u_i \in [0, a_i + 2\delta]$ such that $x_i^\alpha = T^\alpha y_i \in X_i^\alpha$; i.e. $x_i^\alpha \in X_i^\alpha$, where $x_i^\alpha \doteq (x_i^1, x_i^2, \ldots, x_i^N)^T$. Let $\bar{u} \doteq (u_1, u_2, \ldots, u_N)^T$ and $(X_i^\alpha)_{\alpha<\gamma}$, $(X_i^\alpha)_{\alpha>\gamma}$ be the Hilbert scale induced by $L_i$ over $X$ and the multiple Hilbert scale induced by $L$ over $X^N$, respectively. Suppose that the regularization parameter $\alpha$ is chosen as

$$\alpha = \alpha(\delta) \doteq c \delta^{\varepsilon} \quad \text{with} \quad \varepsilon \in \left(0, \min_{1 \leq i \leq N} \left\{ \frac{2(a_i + s_i)}{a_i} \right\} \right),$$  

(66)

where $c > 0$ and, for each index $i$, $1 \leq i \leq N$, $a_i$ is the constant in (64). Then:

(i) $\lim_{\delta \rightarrow 0} x^\alpha_{\delta(\delta)} = x^\alpha$ and, moreover,

(ii) for the total error, the following order of convergence holds:

$$\|x^\alpha_{\delta} - x^\alpha\| = O(\delta^\sigma)$$  

where $\sigma = \min_{1 \leq i \leq N} \min \left\{ 1 - \frac{a_i}{2(a_i + s_i)}, \frac{a_i}{2(a_i + s_i)} \right\}$.

(iii) the order of convergence of the total error in (ii) is optimal when in (66) the value of $\varepsilon$ is chosen as

$$\varepsilon = \left( \max_{1 \leq i \leq N} \frac{a_i}{2(a_i + s_i)} + \min_{1 \leq i \leq N} \frac{u_i}{2(a_i + s_i)} \right)^{-1},$$

in which case $\|x^\alpha_{\delta, \varepsilon} - x^\alpha\| = O(\delta^{\sigma_\varepsilon})$, where $\sigma_\varepsilon \doteq \min_{1 \leq i \leq N} \frac{a_i}{2(a_i + s_i)}$. 

Proof. Applying theorem 3.1 to each operator $L_i$, $1 \leq i \leq N$, since $\varepsilon \leq \frac{2(a_i + s_i)}{a_i}$, with the choice of $\alpha$ as in (66) it follows that

$$\|x^\alpha_{\delta, \varepsilon} - x^\alpha\| = O(\delta^{\sigma_\varepsilon}),$$  

(67)

where

$$x^\alpha_{\delta, \varepsilon} \doteq L_i^{-\frac{\varepsilon}{a_i}} g_\alpha(B_i^T B_i) B_i^T y^\delta$$  

and $\sigma_i = \min \left\{ 1 - \frac{a_i \varepsilon}{2(a_i + s_i)}, \frac{u_i \varepsilon}{2(a_i + s_i)} \right\}$.  

(68)
Then,
\[
\| x_{a}^{\delta, \beta} - x^\dagger \| = \| \tilde{y} (\tilde{L}^* g_a (\tilde{B}^* \tilde{B}) \tilde{B}^\dagger y^\delta - x^\dagger) \| \quad (\text{by (65)})
\]
\[
= \sum_{i=1}^{N} \eta_i \delta^{i} x_{i,a} - x^\dagger \quad (\text{by (63)})
\]
\[
= \sum_{i=1}^{N} \eta_i (x_{i,a}^{\delta} - x^\dagger) \quad \left( \text{since } \sum_{i=1}^{N} \eta_i = 1 \right)
\]
\[
\leq \sum_{i=1}^{N} \eta_i \| x_{i,a}^{\delta} - x^\dagger \|
\]
\[
= \sum_{i=1}^{N} \eta c_i \delta^\eta \quad (\text{by (67)})
\]
\[
\leq C \delta^\eta.
\]

where \( C \) is a positive constant (for instance, for \( \delta \in [0,1] \), \( C \) can be chosen as \( C = \max_{1 \leq i \leq N} c_i \)).

This proves (i) and (ii). To prove (iii), note that from theorem 3.1, more precisely from (30), there exist positive constants \( c_i, d_i, 1 \leq i \leq N \), such that

\[
\| x_{a}^{\delta, \beta} - x^\dagger \| \leq \sum_{i=1}^{N} \eta_i \| x_{i,a}^{\delta} - x^\dagger \|
\]
\[
\leq \sum_{i=1}^{N} \eta_i \left( c_i \delta \alpha^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}} + d_i \alpha^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}} \right) \quad (\text{by (30)})
\]
\[
\leq C_i \delta \alpha^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}} + C_2 \alpha^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}}
\]
\[
= \hat{C}_i \delta^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}} + \hat{C}_2 \delta^{\frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}}}
\]
\[
(\text{by (66)), (69)}.
\]

where \( C_i, \hat{C}_i \) denote generic positive constants.

Finally, from (69) it follows that the order of convergence of the total error is optimal when \( \epsilon \) satisfies \( 1 - \max_{1 \leq i \leq N} \frac{\alpha}{\min_{1 \leq i \leq N} \frac{k_i}{n_i} + \frac{k_0}{n_i}} = \min_{1 \leq i \leq N} \frac{k_i}{n_i} \), that is, when \( \epsilon \) is chosen as \( \epsilon = \left( \max_{1 \leq i \leq N} \frac{k_i}{n_i} + \min_{1 \leq i \leq N} \frac{k_0}{n_i} \right)^{-1} \) in which case, also from (69), it follows that

\[
\| x_{a}^{\delta, \beta} - x^\dagger \| = O(\delta^{\sigma_i}), \quad \text{where } \sigma_i \text{ is given by } \sigma_i = \frac{\min_{1 \leq i \leq N} n_i}{\min_{1 \leq i \leq N} \frac{n_i}{k_i} + \max_{1 \leq i \leq N} \frac{n_i}{k_0}}.
\]

\( \Box \)

**Remark 4.4.** From (63) it follows that the regularized solution \( x_{a}^{\delta, \beta} \) is given by \( \tilde{y} \left( \tilde{L}^* g_a (\tilde{B}^* \tilde{B}) \tilde{B}^\dagger y^\delta \right) \) defined in (65) can also be written in the form \( x_{a}^{\delta, \beta} = \tilde{y} \cdot x_{a}^{\delta} \) where \( x_{a}^{\delta} \in \mathcal{X}^{\delta} \). It is defined by \( x_{a}^{\delta} = (x_{1,a}^{\delta}, x_{2,a}^{\delta}, \ldots, x_{N,a}^{\delta})^T \) with \( x_{i,a}^{\delta} = L_{a}^{-\delta} g_a (B^*_a B_a) B_{i,a} y^\delta \in \mathcal{X}^{\delta} \), \( i = 1, 2, \ldots, N \). It is important to note however that in contrast with what happens in the case \( N = 1 \), where it is known that the regularized solution is in \( \mathcal{X}^{\delta} \), the exact degree of regularity of \( x_{a}^{\delta} \) is not explicitly known since, in general, the \( N \) Hilbert scales \( \mathcal{X}^{\delta} \), \( i = 1, 2, \ldots, N \), are not necessarily related in any way.
4.3. Regularization in multiple Hilbert scales with multiple observations

Note that in theorem 4.3, given a single noisy observation \( y^\delta \) we generated the ‘observation vector’ \( \bar{y}^\delta \in \mathcal{Y}^N \) by using \( N \) copies of \( y^\delta \). In practice, it may happen that \( N \) different observations of \( y \), say \( y^\delta_1, y^\delta_2, \ldots, y^\delta_N \), such that \( \| y^\delta_i - y^\delta \| \leq \delta \forall i = 1, 2, \ldots, N \), be available. In such a case we can use them to construct an observation vector in the form \( \bar{y}^\delta \doteq (y^\delta_1, y^\delta_2, \ldots, y^\delta_N)^T \in \mathcal{Y}^N \). Defining now

\[
x_u^{\bar{y}^\delta} \doteq \bar{y}^\delta \cdot (\mathcal{L}^{\bar{y}^\delta} g_u (\bar{B}^* \bar{B}) \bar{y}^\delta)
\]

(with \( \bar{y}, \bar{s}, g_u, \bar{B}, \alpha = c \delta^\alpha \) as in theorem 4.3), it can be easily seen that the same results of theorem 4.3 remain true. In particular, we have that \( \lim_{\alpha \to 0} \| x_u^{\bar{y}^\delta} - x^\delta \| = 0 \) and \( \| x_u^{\bar{y}^\delta} - x^\delta \| = O(\delta^\alpha) \), where \( \alpha = \min_{\delta \in \mathbb{R}} \min \left\{ 1 - \frac{\delta}{\lambda_{\alpha} + \lambda_{\alpha}^{-1}}, \frac{\delta}{\lambda_{\alpha} + \lambda_{\alpha}^{-1}} \right\} \). However, in this case of regularization in multiple Hilbert scales with multiple observations, it is also possible to utilize different types of regularization methods (i.e. different families of functions \( \{g_u\} \)) for each of the observations \( y^\delta_i \), \( 1 \leq i \leq N \), maintaining the convergence to the exact solution and even improving the order of convergence. This may be of particular interest when certain a priori knowledge about the \( i \)th observation suggests the use of a certain type of regularization method. In order to proceed with the formalization and presentation of this result, we will beforehand need to extend the definition of a ‘function of a self-adjoint operator’ \( f(A) \) to the case in which \( \bar{f} : \mathbb{R} \to \mathbb{R}^N \) is a vector-valued function and \( A \) is a self-adjoint operator in a product space \( \mathcal{X} = \bigotimes_{i=1}^N \mathcal{X}_i \), where \( \mathcal{X}_i \) is a Hilbert space for every \( i = 1, 2, \ldots, N \). Let \( \bar{f} : \mathbb{R} \to \mathbb{R}^N \), \( \bar{f} = (f_1, f_2, \ldots, f_n)^T \), \( \bar{f} \) piecewise continuous, \( \{E^A_i\}_{i \in \mathbb{R}} \) the spectral family of \( A, E^A_i : \mathcal{X} \to \mathcal{X}, E^A_i = (E^A_i, E^A_{i+}, \ldots, E^A_n)^T, E^A_i : \mathcal{X} \to \mathcal{X}_i \) (note that \( E^A_i \) is the \( i \)th component of the projection operator \( E^A_i \) on \( \mathcal{X}_i \)). We define the operator \( \bar{f}(A) \) as the spectral vector-valued integral

\[
\bar{f}(A)\bar{x} = \int_{-\infty}^{\infty} \bar{f}(\lambda) \odot dE^A_i \bar{x} = \begin{pmatrix} \vdots \\ \int_{-\infty}^{\infty} f_i(\lambda) \ dE^A_i \bar{x} \\ \vdots \end{pmatrix}, \tag{70}
\]

where ‘\( \odot \)’ denotes the Hadamard product, with the domain given by

\[
\mathcal{D}(\bar{f}(A)) = \left\{ \bar{x} \in \mathcal{X} : \sum_{i=1}^{N} \int_{-\infty}^{\infty} f_i^2(\lambda) \ d\|E^A_i \bar{x}\|^2 < \infty \right\}.
\]

It is important to note in (70) that in the integral \( \int_{-\infty}^{\infty} f_i(\lambda) \ dE^A_i \bar{x} \), the family \( \{E^A_i\}_{i \in \mathbb{R}} \) is not a spectral family (in fact it is not a partition of unity but rather a parametric family of canonical projections of a spectral family on the product space \( \mathcal{X} = \bigotimes_{i=1}^N \mathcal{X}_i \) onto \( \mathcal{X}_i \); more precisely \( E^A_i = \Pi_i E^A_i \)). However, under the hypothesis of piecewise continuity of \( \bar{f} \), it can be easily seen that the existence of each one of the integrals on the RHS of (70) is guaranteed by the classical theory functional calculus. In fact, given any \( i, 1 \leq i \leq N \), by defining \( \bar{g} : \mathbb{R} \to \mathbb{R}^N \) as \( \bar{g}(\lambda) = (f_1(\lambda), f_2(\lambda), \ldots, f_N(\lambda))^T \), since \( \bar{f} \) is piecewise continuous, so is \( \bar{g} \) and therefore the operator \( \bar{g}(A) \) is well defined and clearly for every \( \bar{x} \in \mathcal{D}(\bar{g}(A)) \) one has that \( \bar{g}(A)\bar{x} = \int_{-\infty}^{\infty} f_i(\lambda) \ dE^A_i \bar{x} \).

With this extension of the concept of a function of a self-adjoint operator to the case of vector-valued functions of self-adjoint operators on product spaces, we are now ready to present the following theorem, which extends the result of theorem 4.3 to the case of multiple observations with vector-valued regularization functions in multiple Hilbert scales.
Theorem 4.5. Let $X, Y, X^N, Y^N, T, \tilde{T}, L, \tilde{L}, (X^N_i)_{i \in \mathbb{N}}, \tilde{s}, B_i = TL_i^{-1}, 1 \leq i \leq N, \tilde{L}, (X^N_i)_{i \in \mathbb{N}}, \tilde{B} = \tilde{T}L_i^{-1}$ and $\tilde{\eta}$, all as in theorem 4.3. For each index $i, 1 \leq i \leq N$, let $g_{a_i} : [0, \|g\|_2^2] \rightarrow \mathbb{R}, a_i > 0$, be a family of piecewise continuous functions and $r_{a_i} (\lambda) = 1 - \lambda g_{a_i} (\lambda)$. Suppose also that each one of the families $\{g_{a_i}\}$ verifies the conditions C1, C2 and C3 of theorem 4.3, that is,

C1: $\forall \lambda \in (0, \|B\|_2^2)$ there holds $\lim_{a_i \rightarrow 0^+} g_{a_i} (\lambda) = \frac{1}{\lambda}$,

C2: $\exists \tilde{c}_i > 0$ such that $\forall \lambda \in (0, \|B\|_2^2)$ and $\forall a_i > 0$ there holds $\|g_{a_i} (\lambda)\| \leq \tilde{c}_i a_i^{-1}$,

C3: $\exists \mu_i > 1$ such that if $\mu \in [0, \mu_i]$ then $\lambda^\mu |r_{\lambda} (\lambda)| \leq \tilde{c}_i a_i^\mu \forall \lambda \in (0, \|B\|_2^2)$,

where the $c_i$ are positive constants. Let us denote now with $\vec{a} = (a_1, a_2, \ldots, a_N)^T$ the ‘vector-valued regularization parameter’ and with $\vec{g}_{\vec{a}} : \mathbb{R} \rightarrow \mathbb{R}^N$ the function given by $\vec{g}_{\vec{a}} (\lambda) = (g_{a_1} (\lambda), g_{a_2} (\lambda), \ldots, g_{a_N} (\lambda))^T$ and let $\vec{g}_{\vec{a}}(\vec{B}^\dagger \vec{B})$ be the linear continuous self-adjoint operator on $X^N$ defined via (70). Let $y \in \mathcal{D}(T^\dagger)$, $y^\dagger, y_1^\dagger, \ldots, y_N^\dagger \in Y$ be such that $\|y_i^\dagger - y\| \leq \delta \forall i = 1, 2, \ldots, N$ and $\delta^\dagger = (y_1^\dagger, y_2^\dagger, \ldots, y_N^\dagger)^T \in Y^N$. We define the regularized solution $x_{\vec{a}}^{\dagger, \vec{g}}$ of problem (1) given the observations $y_1^\dagger, y_2^\dagger, \ldots, y_N^\dagger$, with regularization methods $g_{a_1} (\cdot), g_{a_2} (\cdot), \ldots, g_{a_N} (\cdot)$, in the Hilbert scales $X^N_1, X^N_2, \ldots, X^N_N$ induced by the operators $L_1, L_2, \ldots, L_N$ over $X$, with the weights $\eta_1, \eta_2, \ldots, \eta_N$, as

$$x_{\vec{a}}^{\dagger, \vec{g}} = x_{\vec{a}}^{\dagger, \vec{g}} (\vec{g}_{\vec{a}}, \vec{L}, y^\dagger, \vec{x}) = \vec{\eta} \cdot (\vec{L}^{-1} \vec{g}_{\vec{a}}(\vec{B}^\dagger \vec{B}) \vec{B}^\dagger \vec{y}^\dagger).$$

(71)

Suppose also that $\forall i, 1 \leq i \leq N$. There exists $u_i \in [0, a_i + 2\delta]$ such that $x_i^\dagger = T^\dagger y_i \in X^N_i$, i.e. $x_i^\dagger \in X^N_i$, where $x_i^\dagger = (x_i^1, x_i^2, \ldots, x_i^N)^T$ and $\vec{a} = (u_1, u_2, \ldots, u_N)^T$. If the vector-valued regularization parameter $\vec{a}$ is chosen in the form

$$\vec{a} (\delta) = (c_1 \delta^\epsilon_1, c_2 \delta^\epsilon_2, \ldots, c_N \delta^\epsilon_N)^T$$

(72)

where $c_i > 0$ and $0 < \epsilon_i < \frac{2a_i + 2\delta}{a_i}$, $1 \leq i \leq N$, then

(i) $\|x_{\vec{a}}^{\dagger, \vec{g}} - x_i^\dagger\| \rightarrow 0$ for $\delta \rightarrow 0$;

(ii) $\|x_{\vec{a}}^{\dagger, \vec{g}} - x_i^\dagger\| = O(\delta^\sigma)$, where $\sigma = \min_{1 \leq i \leq N} \sigma_i, \sigma_i = \min \{1 - \frac{a_i}{2(a_i + 2\delta)}, \frac{a_i}{2(a_i + 2\delta)}\};$

(iii) the order of convergence of the total error is optimal when the vector regularization parameter in (72) is chosen such that $\epsilon_i = \frac{2a_i + 2\delta}{a_i}$, in which case one obtains $\|x_{\vec{a}}^{\dagger, \vec{g}} - x_i^\dagger\| = O(\delta^\sigma)$, where $\sigma = \min_{1 \leq i \leq N} \frac{a_i}{a_i + 2\delta};$

(iv) the optimal order $O(\delta^\sigma)$ in (iii) which is obtained with this vector-valued (regularization method) $\vec{g}_{\vec{a}}(\vec{B}^\dagger \vec{B})x$ is at least as good as the optimal order $O(\delta^\sigma)$ which is obtained with a single observation and a scalar $g_{a_i} (\lambda)$ (see theorem 4.3 (iii)).

Proof. Let $\{E_{a_i}^{\delta, \vec{g}}\}_{a_i \in \mathbb{R}}$ and $\{E_{a_i}^{\sigma, \vec{g}}\}_{a_i \in \mathbb{R}}$ denote the spectral families of the operators $\vec{B}^\dagger \vec{B}$ and $B_i^\dagger B_i$, respectively. From the definition of $\vec{B}$ and (60), it can be immediately seen that $[\vec{B}^\dagger \vec{B}] \vec{x}_i = B_i^\dagger B_i \vec{x}_i$ and $[E_{a_i}^{\delta, \vec{g}} \vec{x}_i] = E_{a_i}^{\delta, \vec{g}} \vec{x}_i$ and therefore, from (70), it follows that

$$\vec{g}_{\vec{a}}(\vec{B}^\dagger \vec{B})\vec{x} = (g_{a_1} (B_1^\dagger B_1) x_1, g_{a_2} (B_2^\dagger B_2) x_2, \ldots, g_{a_N} (B_N^\dagger B_N) x_N)^T.$$ 

As in theorem 4.3, let $x_{\vec{a}}^{\dagger, \vec{g}} \in X^N_1$ be defined by

$$x_{\vec{a}}^{\dagger, \vec{g}} = L_i^{-1} g_{a_i} (B_i^\dagger B_i) B_i y_i^{\dagger}. $$

(73)
For each index $i$, $1 \leq i \leq N$, let $\sigma_i = \min\left\{1 - \frac{a_i}{\log(N+1)}, \frac{u_i}{\log(N+1)}\right\}$ and $\sigma = \min_{1 \leq i \leq N} \sigma_i$. Then,

$$\|x_{\sigma} - x^\dagger\| = \left\|\eta_i \left( \hat{L}^{-1} g_{\alpha} (B_{i}^\dagger B_{i}^\top)^{-1} y_{i} - x^\dagger \right) \right\|$$

$$= \left\| \sum_{i=1}^{N} \eta_i \left( \hat{L}^{-1} g_{\alpha} (B_{i}^\dagger B_{i}^\top)^{-1} y_{i} - x^\dagger \right) \right\|$$

$$= \left\| \sum_{i=1}^{N} \eta_i (x_{\sigma}^i - x^\dagger) \right\| \quad \text{(by (73))}$$

$$= \left\| \sum_{i=1}^{N} \eta_i (x_{\sigma}^i - x^\dagger) \right\| \quad \text{(since $\sum_{i=1}^{N} \eta_i = 1$)}$$

$$\leq \sum_{i=1}^{N} \eta_i \left\| x_{\sigma}^i - x^\dagger \right\|$$

$$\leq C \delta^{\sigma^*} \quad \text{(for $\sigma^*$ as in (72), by theorem 3.1 (ii))}$$

This proves (i) and (ii).

Now, if the vector-valued regularization parameter $\hat{a}$ in (72) is chosen so that $e_i = \frac{2(a_i + s_i)}{a_i + s_i}$, $\forall i = 1, 2, \ldots, N$, then by virtue of theorem 3.1 (iii) it follows that there exist positive constants $c_1, c_2, \ldots, c_N$, such that $\|x_{\sigma}^i - x^\dagger\| \leq c_i \delta^{\sigma^*}$, $\forall i = 1, 2, \ldots, N$. Then, from (74) we obtain that

$$\|x_{\sigma} - x^\dagger\| \leq \sum_{i=1}^{N} \eta_i c_i \delta^{\sigma^*}$$

$$\leq C \delta^{\sigma^*}$$

where $\sigma^* = \min_{i \in \mathbb{N}} \frac{a_i}{a_i + u_i}$. It is also clear that for $u_i$ and $a_i$ fixed, this order of convergence is optimal and, as we can see, independent of the choice of $\delta$. This proves (iii).

Finally, to prove (iv) we must verify that $\sigma_i \leq \sigma^*$, where $\sigma_i$ is the optimal order in theorem 4.3 (iii), that is,

$$\sigma_i = \min_{i \in \mathbb{N}} \frac{u_i}{2(a_i + s_i)}.$$

For that, observe that since $a_i$, $u_i$ and $s_i$ are all positive, there holds

$$\max_{i \in \mathbb{N}} \left( \frac{2(a_i + s_i)}{u_i} \right) \max_{i \in \mathbb{N}} \left( \frac{a_i}{2(a_i + s_i)} \right) \geq \max_{i \in \mathbb{N}} \left( \frac{a_i}{u_i} \right),$$

or equivalently

$$\max_{i \in \mathbb{N}} \left( \frac{a_i}{2(a_i + s_i)} \right) \geq \max_{i \in \mathbb{N}} \left( \frac{a_i}{u_i} \right),$$

from where it follows that

$$\frac{1}{1 + \max_{i \in \mathbb{N}} \left( \frac{a_i}{u_i} \right)} \leq \frac{1}{1 + \max_{i \in \mathbb{N}} \left( \frac{a_i}{u_i} \right)},$$

$$\leq C \delta^{\sigma^*}$$.
\[ \begin{align*}
\| x_{\alpha}^\delta - x^* \| &= O(\delta),
\end{align*} \]

where \( \delta \) is chosen in the form
\[ \delta = (c_1 \delta_1^{-\frac{1}{4m+1}}, c_2 \delta_2^{-\frac{1}{2m+1}}, \ldots, c_N \delta_N^{-\frac{1}{2m+1}}) \],

and therefore
\[ \min_{1 \leq i \leq N} \frac{u_i}{a_i + u_i} \leq \min_{1 \leq i \leq N} \left( \frac{u_i}{a_i} \right), \]

that is, \( \sigma_i \leq \sigma^* \), as we wanted to prove.

Here again note that the regularized solution \( x_{\alpha}^\delta \) given by (81) can be written as \( x_{\alpha}^\delta = \tilde{\eta} \cdot x_{\alpha}^\delta \)
where \( x_{\alpha}^\delta = L^{-i} g_{\tilde{\alpha}}(B^* B) y^\delta \in X_{\alpha}^* \).

In the presence of a fixed noise level \( \delta \) in the \( N \) observations \( y_1^\delta, y_2^\delta, \ldots, y_N^\delta \), in light of theorem 3.1 (iii), one should not expect that the order of convergence \( O(\delta^\sigma) = O(\delta \min_{1 \leq i \leq N} \frac{u_i}{a_i + u_i}) \) in theorem 4.5 can be improved. However, if the noise levels can be controlled, then by appropriately handling them on those components on which it is known that the degree of regularity of the exact solution \( x^* \) on the corresponding Hilbert scale (measured in terms of \( a_i \)) is relatively small or the corresponding parameter of comparison of relative regularity between the operators \( T \) and \( L_{\alpha}^{-1} \), measured in terms of \( a_i \) (see (64)), is relatively large, then the order of convergence \( O(\delta^\sigma) \) cannot be fact improved. More precisely we have the following result.

**Theorem 4.6.** Let \( \mathcal{X}, \mathcal{Y}, \mathcal{X}^N, \mathcal{Y}^N, T, \tilde{T}, L_i, u_i, a_i, 1 \leq i \leq N, \tilde{u}, \tilde{s}, \tilde{L}, \tilde{B} = \tilde{T} L^{-i}, B_i = T L_i^{-1}, \tilde{g}_i, \tilde{y} \in D(T^i), x_i \in X_{\alpha}^i \) and \( \hat{\eta} \), all as in theorem 4.5. Let \( y_1^\delta, y_2^\delta, \ldots, y_N^\delta \in \mathcal{Y} \) be such that \( \| y_i^\delta - y \| \leq \delta_i \forall i = 1, 2, \ldots, N, \delta = (\delta_1, \delta_2, \ldots, \delta_N)^T \) and \( y^\delta = (y_1^\delta, y_2^\delta, \ldots, y_N^\delta)^T \in \mathcal{Y}^N \) and define now the regularized solution \( x_{\alpha}^\delta \) of the problem \( T x = y \) as

\[ \hat{x}_{\alpha}^\delta = x_{\alpha}^\delta \tilde{g}_i, \tilde{L}, \tilde{y}^\delta, \tilde{s} \tilde{\eta}, x^\delta \]

where \( \tilde{x}_{\alpha}^\delta = L^{-1} g_i(B^* B) y^\delta \in X_{\alpha}^* \). If \( \delta_i = \delta_i(\delta) = \delta^0 \) with
\[ \delta_0 = \max_{1 \leq i \leq N} \frac{u_i}{a_i + u_i}, \]

for every \( 1 \leq i \leq N \), and the vector-valued regularization parameter \( \tilde{\alpha}(\delta) \) is chosen in the form
\[ \tilde{\alpha}(\delta) = (c_1 \delta_1^{-\frac{1}{4m+1}}, c_2 \delta_2^{-\frac{1}{2m+1}}, \ldots, c_N \delta_N^{-\frac{1}{2m+1}})^T, \]

where \( c_1, c_2, \ldots, c_N \) are arbitrary positive constants, then
\[ \| x_{\alpha}^\delta - x^* \| = O(\delta), \]

where \( \delta \) is chosen in the form
\[ \delta = (\delta_1^{-\frac{1}{4m+1}}, \delta_2^{-\frac{1}{2m+1}}, \ldots, \delta_N^{-\frac{1}{2m+1}})^T, \]

**Proof.** Let \( \alpha_i = c_i \delta_i^{-\frac{1}{4m+1}} \) and \( x_{\alpha_i}^\delta = L_i^{-1} g_i(B^* B) y_i^\delta \). By virtue of theorem 3.1 (iii), it follows that there exist constants \( k_1, k_2, \ldots, k_N \) such that
\[ \| x_{\alpha_i}^\delta - x_i^* \| \leq k_i \delta_i^{-\frac{1}{4m+1}}, 1 \leq i \leq N. \]
On the other hand, by following the same steps as in theorem 4.5, for $\tilde{x}_{\delta}^\theta$ defined as in (75) one has that

$$\|\tilde{x}_{\delta}^\theta - x^\dagger\| \leq \sum_{i=1}^{N} \eta_i \|x_{i,\delta}^\theta - x_i^\dagger\|$$

$$\leq \sum_{i=1}^{N} \eta_i \delta_i^{\frac{\alpha_i}{\max_{1\leq i\leq N} \alpha_i}} \quad \text{(by (79))}$$

$$= \sum_{i=1}^{N} \eta_i \delta^{\frac{\alpha_i}{\max_{1\leq i\leq N} \alpha_i}} \quad \text{(since $\delta_i = \delta^p$)}$$

$$\leq C_0 \delta^{\frac{\alpha}{\max_{1\leq i\leq N} \alpha_i}} \quad \text{(by (76))}$$

$$= C_0 \delta^\theta.$$  \[ \square \]

Note that in order to obtain the order of convergence in (78), it is necessary that the noise level in the $i$th component be $\delta_i = \delta^p$ with $p_i \geq \frac{\max_{1\leq i\leq N} \alpha_i}{\min_{1\leq i\leq N} \alpha_i} \quad (\geq 1 \forall i)$. Hence, the precision in the observation measurements must be improved precisely in those components for which the regularity of $x_i^\dagger$ as an element of the corresponding Hilbert scale, measured by $u_i$, is relatively small or the parameter $a_i$ is large.

5. Potential applications and open issues

During the last decade, Hilbert scales have been successfully applied to several concrete problems. Among them we mention the reconstruction of reaction and diffusion coefficients and identification of parameters in PDEs [10, 4], solution of Fredholm, Abel and Hammerstein integral equations, applications in computerized tomography [4, 5], etc.

Although the central purpose of this paper is strongly theoretical and mainly oriented to introducing the theory of multiple Hilbert scales, as well as the convergence results and error estimates presented in section 4, we briefly describe here some tips for potential applications of this theory as well as some open problems and issues which we believe deserve further research. Some of these points have been already been briefly mentioned or touched upon in the course of this paper.

In article [16], general conditions guaranteeing the existence, uniqueness and stability of minimizers of generalized Tikhonov–Phillips functionals of the form $J_{\alpha, W}(x) = \|Tx - y\|^2 + \alpha W(x)$ were obtained. These conditions were studied in particular for penalizers $W(\cdot)$ of the form $W(x) = ||Lx||^2$ and $W(x) = \sum_{i=1}^{N} \alpha_i ||L_ix||^2$, among others. Also the convergence of the corresponding regularized solutions for the case of differentiable parameter choice rules was proved in [15]. For both choices of penalizers, it is clear that the knowledge that the exact solution $x^\dagger$ belongs to $D(L)$ or $\bigcap_{1\leq i \leq p} D(L_i)$, respectively, is required. It is precisely in the absence of this type of information or knowledge that Hilbert scales can play a very important role for the solution of the corresponding inverse problem. At this point, it is timely to mention that the penalizer $W(\cdot)$ in the functional $J_{\alpha, W}$ is used not only to stabilize the inversion of the ill-posed problem but also to enforce certain characteristics of the approximating solutions and of the particular limiting least-squares solution that they approximate. Hence, it is reasonable to assume that an adequate choice of penalizer, based on a priori knowledge of certain characteristics of the exact solution of the problem at hand, will lead to approximated regularized solutions which will appropriately reflect those characteristics. Similarly, it is
reasonable to think that the simultaneous use of two or more penalizers of a different nature
will allow, in some way, different characteristics of the exact solution to be captured by the
approximated regularized solution. This is of particular interest, for instance, in certain image
restoration problems where it is known a priori that the original image possesses both very
irregular and very irregular regions, perhaps with pronounced borders and discontinuities.

Such is the case for example of ‘blocky’ images frequently appearing in medicine, in the
detection of tumoral regions, and also in astronomy. It was precisely the idea of combining
the advantages of the use of this type of general penalizer [16] with the adaptive virtues that
Hilbert scales possess (in regard to the convergence of the total error as a function of the
degree of regularity of the exact solution), that motivated the theory of multiple Hilbert scales
developed in this paper. It is important to point out, however, that the theory of multiple Hilbert
scales is potentially applicable to a much broader set of problems.

In a concrete practical problem, one or more degraded noisy observations could be
available. In the case of a single observation, convergence of a certain order of the total error
theorem 2.10 is immediately guaranteed. This order can be improved if multiple observations
are available. In this case, the simultaneous use of different types of regularization methods (i.e.
different families of functions \{g_\alpha\}) could result in capturing different characteristics of the
exact solution. This could also be of particular interest when certain a priori knowledge about
the i-th observation (such as the type of blurring or noise) suggests the use of a
certain type of regularization in the i-th component of the observation vector. Clearly, when
\hat{g}_\alpha is chosen such that \hat{g}_\alpha(\lambda) = \frac{1}{\lambda + \delta}, the regularized solutions in multiple Hilbert scales
defined in (75) coincide with the minimizers of the generalized Tikhonov–Phillips functional

\[ J_\alpha(x) = \|Tx - y\|^2 + \sum_{\alpha=1}^{N} \alpha_i \|L_\alpha x\|^2. \]

The use of multiple Hilbert scales could also be of particular interest in the case of multiple
observations where some control on the noise level is available. As shown in theorem 4.6, the
order of convergence of the exact solution obtained for a fixed noise level can be improved
and even optimized by appropriately handling the noise levels, by reducing them in those
components for which the regularity of \(x^\delta\) as an element of the corresponding Hilbert scale,
measured by \(u_i\), is relatively small or the comparison parameter \(a_i\) is large.

Finally we discuss below some open problems and issues which we believe deserve further
study.

First we recall that within the standard theory of (single) Hilbert scales it is a well
known fact that the regularized solution defined in (21) belongs to \(X^\delta = \mathcal{D}(L^\delta)\). Hence,
once \(s\) is chosen, the smoothness of the regularized solution is explicitly known in terms of the operator \(L\). However, as noted in sections 4.2 and 4.3, a regularized solution in the context of the present theory of multiple Hilbert scales can be written
in the form \(x^{\delta,\eta}_a = x^{\delta,\eta}_a(\hat{g}_\alpha, L, \eta^\delta, \hat{s}) \approx \hat{\eta}_s x^{\delta,\eta}_a\), where \(\hat{\eta}\) is a vector of weights and
\(x^{\delta,\eta}_a \in L^{-\delta} \hat{g}_\alpha(B^\delta B^\delta)^{-1} = \begin{pmatrix} \hat{\alpha}_1 a_1, \ldots, \hat{\alpha}_N a_N \end{pmatrix}^T \in X^\delta_t\). Thus, \(x^{\delta,\eta}_a\) is guaranteed to have
at least the minimum smoothness of the \(x^{\delta,\eta}_a\) as elements of \(X^\delta_{\eta,\alpha}\). However, in general, the
exact degree of regularity will not be explicitly known since the \(N\) Hilbert scales \((X^\delta_{\eta,\alpha})_{\alpha \in \mathbb{R}}\) are not necessarily related in any way. It would be interesting to find general conditions on the operators \(L_a\), and perhaps on \(T\), allowing the explicit and a priori knowledge of a minimum or
the exact degree of regularity of \(x^{\delta,\eta}_a\).

An interesting open problem, already pointed out in [15] in a very particular setting, is
that of appropriately choosing the weight vector \(\hat{\eta}\).

For the case of single Hilbert scales, Neubauer extended the Natterer’s original results,
obtaining convergence rates for the case in which the regularization parameter is chosen by
means of Morozov’s discrepancy principle (see [18]). We note however that all of the parameter
In this paper several convergence results in Hilbert scales under different source conditions were proved and orders of convergence and optimal orders of convergence were derived. Also relations between those source conditions were proved. The concept of a multiple Hilbert scale on a product space was introduced, and regularization methods on these scales were defined, first for the case of a single observation and then for the case of multiple observations. In the latter case, it was shown how vector-valued regularization functions in these multiple Hilbert scales can be used. In all cases, convergence was proved and orders and optimal orders of convergence were shown. Finally, some potential applications and open problems were discussed.

Acknowledgments

This work was supported in part by Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, through PIP 2010–2012 Nro. 0219, by Universidad Nacional del Litoral, U.N.L., through project CAI+D 2009-PI-62-315, by Agencia Nacional de Promoción Científica y Tecnología ANPCyT, through project PICT 2008-1301 and by the Air Force Office of Scientific Research, AFOSR, through grant FA9550-10-1-0018.

References

[1] Acar R and Vogel C R 1994 Analysis of bounded variation penalty methods for ill-posed problems Inverse Problems 10 1217–29
[2] Dautray R and Lions J-L 1990 Mathematical Analysis and Numerical Methods for Science and Technology, Vol. 3: Spectral Theory and Applications (Berlin: Springer)
[3] Egger H 2005 Preconditioning iterative regularization methods in Hilbert scales PhD Thesis Johannes Kepler Universität
[4] Egger H 2006 Semi-iterative regularization in Hilbert scales SIAM J. Numer. Anal. 44 66–81
[5] Egger H and Neubauer A 2005 Preconditioning Landweber iteration in Hilbert scales Numer. Math. 101 643–62
[6] Engl H W, Hanke M and Neubauer A 1996 Regularization of Inverse Problems (Mathematics and Its Applications vol 375) (Dordrecht: Kluwer)
[7] Friedrichs K O 1973 Spectral Theory of Operators in Hilbert Space (Applied Mathematical Sciences vol 9) (New York: Springer)
[8] Hadamard J 1902 Sur les problèmes aux dérivées partielles et leur signification physique Princeton Univ. Bull. 13 49–52
[9] Heinz E 1951 Beiträge zur Störungstheorie der Spektralzerlegung Math. Ann. 123 415–38
[10] Hohage T and Pricop M 2008 Nonlinear Tikhonov regularization in Hilbert scales for inverse boundary value problems with random noise Inverse Problems Imag. 2 271–90
[11] Jin Q and Tautenhahn U 2011 Implicit iteration methods in Hilbert scales under general smoothness conditions
Inverse Problems 27 1–27
[12] Kohler J and Tautenhahn U 1995 Error bounds for regularized solutions of nonlinear ill-posed problems
Inverse Ill-Posed Problems 3 47–74
[13] Krein S G and Petunin Ju I 1966 Scales of Banach spaces Uspehi Mat. Nauk 21 89–168
[14] Lin F and Nashed M Z 1997 Tikhonov regularization of nonlinear ill-posed problems with closed operators in
Hilbert scales J. Inverse Ill-Posed Problems 5 363–76
[15] Mazzieri G L and Spies R D 2011 Radial and differentiable-rule convergence for Tikhonov–Phillips methods
with penalizers given by linear combination of seminorms induced by closed operators (submitted)
[16] Mazzieri G L, Spies R D and Temperini K G 2011 Existence, uniqueness and stability of solutions of generalized
Tikhonov–Phillips functionals IMA Preprint Series no 2375 Institute for Mathematics and Its Applications,
University of Minnesota (arXiv:1108.4067v1)
[17] Natterer F 1984 Error bounds for Tikhonov regularization in Hilbert scales Appl. Anal. 18 29–37
[18] Neubauer A 1988 An a posteriori parameter choice for Tikhonov regularization in Hilbert scales leading to
optimal convergence rates SIAM J. Numer. Anal. 25 1313–26
[19] Neubauer A 1992 Tikhonov regularization for nonlinear ill-posed problems in Hilbert scales Appl. Anal. 46 59–72
[20] Pazy A 1983 Semigroups of Linear Operators and Applications to Partial Differential Equations (Applied
Mathematical Sciences vol 44) (New York: Springer)
[21] Phillips D L 1962 A technique for the numerical solution of certain integral equations of the first kind J. Assoc.
Comput. Mach. 9 84–97
[22] Schröter T and Tautenhahn U 1994 Error estimates for Tikhonov regularization in Hilbert scales Numer. Funct.
Anal. Optim. 15 155–68
[23] Seidman T I 1980 Nonconvergence results for the application of least-squares estimation to ill-posed problems
J. Optim. Theory Appl. 30 535–47
[24] Spies R D and Temperini K G 2006 Arbitrary divergence speed of the least-squares method in infinite-
dimensional inverse ill-posed problems Inverse Problems 22 611–26
[25] Tautenhahn U 1994 Error estimates for regularized solutions of nonlinear ill-posed problems Inverse
Problems 10 485–500
[26] Tautenhahn U 1998 On a general regularization scheme for nonlinear ill-posed problems: II. Regularization in
Hilbert scales Inverse Problems 14 1607–16
[27] Tikhonov A N 1963 Regularization of incorrectly posed problems Sov. Math. Dokl. 4 1624–7
[28] Tikhonov A N 1963 Solution of incorrectly formulated problems and the regularization method Sov. Math. Dokl.
4 1035–8
[29] Whaba G 1990 Spline Models for Observational Data (Philadelphia, PA: SIAM)