Discretization strategy for ill-posed problems in variable Hilbert scales

Peter Mathé† and Sergei V. Pereverzev‡
† Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, D–10117 Berlin, Germany
E-mail: mathe@wias-berlin.de
‡ National Academy of Sciences of Ukraine
Institute of Mathematics
Tereshenkovska Str. 3
Kiev 4
Ukraine
E-mail: serg-p@mail.kar.net

AMS classification scheme numbers: 65J20

Submitted to: Inverse Problems

Abstract. The authors study the regularization of projection methods for solving linear ill-posed problems with compact and injective linear operators in Hilbert spaces. Smoothness of the unknown solution is given in terms of general source conditions, such that the framework of variable Hilbert scales is suitable.

The structure of the error is analyzed in terms of the noise level, the regularization parameter and as a function of other parameters, driving the discretization. As a result, a strategy is proposed, which automatically adapts to the unknown source condition, uniformly for certain classes, and provides the optimal order of accuracy.
1. Introduction

The present study continues research, started by the authors in [7]. It extends the analysis of ill-posed problems in variable Hilbert scales towards the issue of regularization of projection methods. We study linear ill-posed equations

$$y_\delta = Ax + \delta \xi,$$

(1)

where \( A: X \to Y \) is a compact injective operator acting between Hilbert spaces, \( \xi \) is uniformly bounded noise, \( \| \xi \| \leq 1 \), and \( \delta \) represents the known noise level. The goal is to reconstruct \( x \) from (discretized) data \( y_{\delta,1}, \ldots, y_{\delta,n} \).

The best possible accuracy, regardless of any discretization is determined by some a priori assumption on the exact solution \( x \). For the case

$$A_\phi := \{ x \in X, \quad x = (A^*A)^{p/2}v, \quad \| v \| \leq 1 \},$$

(2)

discretization was studied in [9]. This setup of a priori smoothness does not allow to treat severely ill-posed problems. Therefore, the framework of variable Hilbert scales, as introduced in [3, 4], turns out to be fruitful. Here, smoothness is more generally given in terms of some index function \( \varphi \) on the spectrum of \( A^*A \) by

$$A_\varphi := \{ x \in X, \quad x = \varphi(A^*A)v, \quad \| v \| \leq 1 \},$$

where \( A_\varphi \) is called source condition. The present study extends the error analysis from [9] to the case of general source conditions. This requires to develop new tools, as there are interpolation in variable Hilbert scales and operator concave functions. For the convenience of the reader, and since these results might be interesting on their own, they are presented in two appendices.

The organization of the material is as follows. First, in §2 we briefly recall some concepts from [7]. In particular we recall the notion of qualification of a regularization in variable Hilbert scales, and what it means, that this covers the given smoothness \( \varphi \). Then, in §3 we regard the discretization of the original operator \( A \) as a perturbation and analyse the error resulting from discretizing the perturbation. This initial analysis closes in §3.3 indicating the basic tasks to be carried out.

The benchmark for any error estimate of a projection method, say \( B := QAP \), with finite projections \( P \) and \( Q \) is the intrinsic best possible accuracy \( e(A_\varphi, \delta) \), which was shown to be of the order \( \delta \to \varphi(\Theta^{-1}(\delta)) \), as \( \delta \to 0 \), under fairly general assumptions, we refer to [3, 4, 11] and also [7]. Above \( \Theta(t) := \sqrt{t}\varphi(t) \). The error estimates as provided in §3 are different under different assumptions on the index function. Thus we will introduce several classes of index functions, below. Nevertheless, it is common for all classes, that the error bounds of a projection scheme \( B = QAP \) depend on the chosen regularization parameter \( \alpha \) and bounds on \( \rho := \| A(I - P) \| \) and on \( \eta := \| (I - Q)A \| \). Thus, in §4 we will propose an adaptation strategy for choosing \( \rho \) and \( \eta \) as functions of \( \alpha \) and the known noise level \( \delta \), to maintain the best possible order of accuracy. This gives rise to a discretized version of adapting the regularization parameter \( \alpha \) to the unknown source condition, presented and analyzed in §4.2. We complete the study with a discussion in §5.
2. Preliminaries

We recall that the ill-posed problem is given by

\[ y_\delta = Ax + \delta \xi. \]

The a priori assumption on the smoothness of \( x \) is given through \( \varphi \) by some source condition [2]. For regularization we shall assume more specifically, that the index function is continuous, increasing and satisfies \( \varphi(0) = 0, \varphi(a) = 1 \), where \( a \) is such that \( \|A\|^2 < a \). Further restrictions will be necessary later on.

We are interested in regularization methods given by some operator function \( \alpha \rightarrow g_\alpha(A^*A) \), \( 0 < \alpha \leq a \), i.e., the approximation to \( x \in A_\varphi \) is given by choosing some \( \alpha = \alpha(\delta) \) and letting

\[ x_{\alpha,\delta} := g_\alpha(A^*A)A^*y_\delta. \]

First we need the notion of the qualification of regularization.

**Definition 1** A family \( g_\alpha \), \( 0 < \alpha \leq a \) is called regularization, if there are constants \( \gamma_\ast \) and \( \gamma \) for which

\[
\sup_{0 < \lambda \leq a} |1 - \lambda g_\alpha(\lambda)| \leq \gamma, \quad 0 < \alpha \leq a.
\]

and

\[
\sup_{0 < \lambda \leq a} \sqrt{\lambda} |g_\alpha(\lambda)| \leq \frac{\gamma_\ast}{\sqrt{\alpha}}, \quad 0 < \alpha \leq a,
\]

The regularization \( g_\alpha \) is said to have qualification \( \rho \), for an increasing function \( \rho : (0, a) \rightarrow \mathbb{R}_+ \), if

\[
\sup_{0 < \lambda \leq a} |1 - \lambda g_\alpha(\lambda)| \rho(\lambda) \leq \gamma \rho(\alpha), \quad 0 < \alpha \leq a.
\]

We are interested in analyzing regularizations, based on the operator \( B^*B \). The error of any regularization method \( y_\delta \rightarrow g_\alpha(B^*B)B^*y_\delta \), is considered uniformly over problem elements from \( A_\varphi \), i.e.,

\[ e(A_\varphi, g_\alpha(B^*B)B^*, \delta) := \sup_{x \in A_\varphi} \sup_{\|\xi\| \leq 1} \|x - g_\alpha(B^*B)B^*y_\delta\|. \]

If \( A = B \), then the error of regularization has been studied in [7]. We recall some basic result from this study.

Given an index function, say \( \varphi \) and regularization with qualification, say \( \rho \), we shall say, that \( \rho \) covers \( \varphi \), if

\[
\frac{\rho(\alpha)}{\varphi(\alpha)} \leq \inf_{0 < \lambda \leq a} \frac{\rho(\lambda)}{\varphi(\lambda)}, \quad 0 < \alpha \leq a.
\]  

(3)

**Remark 1** In [7], a slightly weaker condition is imposed on the interplay between the index function and the qualification, namely, that (3) is fulfilled up to some constant, say \( 0 < c \leq 1 \). Although this weaker notion cannot be obtained by rescaling, all analysis below remains true, by simply replacing \( \gamma \) by \( \gamma/c \).

The following result has been proven in [7, Thm. 2].
Theorem 1 Let $\varphi$ be any increasing index function and let $\bar{\alpha}$ be chosen to satisfy
\[ \sqrt{\bar{\alpha}} \varphi(\bar{\alpha}) = \delta. \]
If $g_\alpha$ is any regularization of qualification that covers $\varphi$, then uniformly for $x \in A_\varphi$ it holds
\[ \|x - g_\alpha(A^*A)A^*y_b\| \leq (\gamma + \gamma_*) \varphi(\Theta^{-1}(\delta)), \quad 0 < \delta \leq a, \]
where $\Theta(t) := \sqrt{t} \varphi(t)$.

Remark 2 It has been shown in [7, Thm. 1], that typically this bound cannot be improved as far as the order of accuracy is concerned. Therefore, the rate given by $\delta \rightarrow \varphi(\Theta^{-1}(\delta))$ serves as a benchmark for any error estimate.

3. Perturbation of ill-posed problems

The objective in this section is the study of equation (1), where the source condition is given in the scale generated by $A^*A$, but which is regularized with an operator $B^*B$, which may be thought of as an approximation to $A^*A$. The question is how close to $A$ the operator $B$ needs to be, such that the corresponding regularized solution achieves the optimal order of accuracy? To deal with the case of discrete data, we shall assume, that data are given through an orthogonal design, corresponding to an orthogonal projection $Q := Q_n$ as
\[ Q_n y_b = Q_n A x + \delta Q_n \xi. \]
The approximate solution is calculated through $B^*B$ as regularization $g_\alpha(B^*B)B^*Q_n y_b$, such that we assume $B^*Q_n = B^*$, that $Q_n B = B$, as typical for projection schemes $B := Q_n A P_m$, for another projection $P_m$. Moreover, we assume, that $g_\alpha$ covers the smoothness $\varphi$. Then we have the following obvious error decomposition at the true solution $x$.
\[ \|x - g_\alpha(B^*B)B^*Q_n y_b\| \leq \|x - g_\alpha(B^*B)B^*Q_n A x\| + \delta \|g_\alpha(B^*B)B^*Q_n \xi\|. \]
The noise term is immediately estimated as
\[ \delta \|g_\alpha(B^*B)B^*Q_n \xi\| \leq \gamma_* \delta / \sqrt{\alpha}. \]
We bound the noise free term as follows
\[
\|x - g_\alpha(B^*B)B^*Q_n A x\| \leq \|x - g_\alpha(B^*B)B^*B x\| + \|g_\alpha(B^*B)B^*(B - Q_n A) x\|
\leq \gamma \varphi(\alpha) + \|((I - g_\alpha(B^*B)B^*B)(\varphi(A^*A) - \varphi(B^*B)))\|
+ \frac{\gamma_*}{\sqrt{\alpha}} \|B - Q_n A x\|. 
\]
We arrive at
\[
\begin{align*}
e(x, g_\alpha(B^*B)B^*, \delta) & \leq \gamma \varphi(\alpha) + \|((I - g_\alpha(B^*B)B^*B)(\varphi(A^*A) - \varphi(B^*B)))\|
+ \frac{\gamma_*}{\sqrt{\alpha}} (\delta + \|B - Q_n A x\|).
\end{align*}
\]
\[ \dagger \text{We will adopt the convention, that operator norms are assumed to be for operators acting from } X \rightarrow X \text{ or } X \rightarrow Y, \text{ which will be clear from the context, if not indicated otherwise.} \]
In light of the best possible order of accuracy without discretization, which can be represented as \( \varphi(\alpha) + \delta/\sqrt{\alpha} \), the discretization has to be chosen, such that

\[
\| (B - Q_n A) : X_\varphi \to Y \| \simeq \delta
\]

and

\[
\| (I - g_\alpha(B^A B^A B^A)(\varphi(A^A) - \varphi(B^A B^A))) \| \simeq \varphi(\alpha) + \delta/\sqrt{\alpha}.
\]

(5)

Here \( X_\varphi \) is a normed space with a unit ball \( [2] \). For more details we refer to Appendix A. The left hand side in (5) can further be rewritten, introducing \( P \) such that \( I - P \) is the orthogonal projection onto the kernel of \( B \), i.e., \( BP = B \) as

\[
\| (I - g_\alpha(B^A B^A B^A))(\varphi(A^A) - \varphi(B^A B^A)) \| = \| (I - g_\alpha(B^A B^A B^A)(P \varphi(A^A) - \varphi(B^A B^A))) \|,
\]

since \( P \varphi(B^A) = \varphi(B^A B^A) \).

Further rendering will depend on whether \( \varphi \) allows a certain norm estimate, or it can be split into a part with this property and a Lipschitz part. Precisely, we require to restrict to the classes \( \mathcal{F}(d) \) and \( \mathcal{F}_L(d) \) of index functions, given as

\[
\mathcal{F}(d) : = \{ \varphi, \quad \| \varphi(A^A) - \varphi(B^A B^A) \| \leq d\varphi(\| A^A - B^A B^A \|) \}
\]

and

\[
\mathcal{F}_L(d) : = \{ \varphi = \vartheta \psi, \quad \psi \in \mathcal{F}(d), \quad \vartheta \text{ incr. and Lipschitz, } \vartheta(0) = 0 \},
\]

which means, that

\[
\| \vartheta(A^A) - \vartheta(B^A B^A) \| \leq d\| A^A - B^A B^A \|.
\]

The splitting into the Lipschitz part \( \vartheta \) and \( \psi \) is not unique, such that we implicitly assume, that the Lipschitz constant is equal to 1.

**Remark 3** As has been observed in [6], functions from class \( \mathcal{F}(d) \) behave like moduli of continuity, such that they cannot converge faster than linearly to 0. To overcome this restriction, the class \( \mathcal{F}_L(d) \) is introduced.

As can be seen from [6,13], functions \( t \to t^\mu \) belong to \( \mathcal{F}(1) \) as long as \( 0 < \mu \leq 1 \), whereas they are (operator) Lipschitz continuous for \( \mu \geq 1 \).

### 3.1. Index functions from class \( \mathcal{F}(d) \)

For source conditions with index functions from \( \mathcal{F}(d) \) we immediately can bound

\[
e(A_\varphi, g_\alpha(B^A B^A B^A), \delta) \leq \gamma(\varphi(\alpha) + \| P \varphi(A^A) - \varphi(B^A B^A) \|) + \frac{\gamma_0}{\sqrt{\alpha}} (\delta + \| (B - Q_n A) : X_\varphi \to Y \|).
\]
3.2. Index functions from class $\mathcal{F}_L(d)$

To treat source conditions with index functions from $\mathcal{F}_L(d)$ we make the following observation. If the qualification of $g_\alpha(B^*B)B^*$ covers $\varphi$, then it also covers $\vartheta$. Using this we can estimate the left hand side in (5) as

$$\| (I - g_\alpha(B^*B)B^*(B^*B)\vartheta(B^*B) - \psi(A^*A)) \|$$

$$\leq \| (I - g_\alpha(B^*B)B^*(B^*B)\vartheta(B^*B) - \psi(A^*A)) \|$$

$$+ \| (I - g_\alpha(B^*B)B^*(B^*B)\vartheta(A^*A) - \vartheta(B^*B)\psi(A^*A)) \|.$$

The first summand on the right can further be treated using the qualification as

$$\| (I - g_\alpha(B^*B)B^*(B^*B)\vartheta(B^*B) - \psi(A^*A)) \| \leq \gamma \vartheta(\alpha) \| P\psi(A^*A) - \psi(B^*B) \|.$$

Using the Lipschitz property of $\vartheta$ we can bound the second summand on the right in (6) as

$$\| (I - g_\alpha(B^*B)B^*(B^*B)\vartheta(A^*A) - \vartheta(B^*B)\psi(A^*A)) \| \leq \gamma \| A^*A - B^*B \|,$$

resulting in the error bound

$$e(A_\varphi, g_\alpha(B^*B)B^*, \delta) \leq \gamma (\varphi(\alpha) + \vartheta(\alpha)) \| P\psi(A^*A) - \psi(B^*B) \| + \| A^*A - B^*B \|$$

$$+ \frac{\gamma_s}{\sqrt{\alpha}} (\delta + \| (B - Q_nA): X_\varphi \to Y \|).$$

This estimate corresponds to the one for functions from class $\mathcal{F}(d)$ above.

3.3. Summary

In order to assess the quality of a perturbed regularization $g_\alpha(B^*B)B^*Q_ny_\delta$ we need to gain information on the following quantities:

$$\| B - Q_nA: X_\varphi \to Y \|;$$

(7)

for any function $\psi$ from $\mathcal{F}(d)$

$$\| P\psi(A^*A) - \psi(B^*B) \|;$$

(8)

Moreover, for functions from $\mathcal{F}_L(d)$ we additionally need to bound

$$\| A^*A - B^*B \|.$$

(9)

The term (7) can be estimated, using some interpolation arguments, as established below in Appendix A. The last term (9) can be treated directly, since no index function is involved there. Crucial work only needs to be done, bounding (8), where we need to rely on techniques concerned with operator monotone functions, as established in [6].

This will be done in Appendix B.

In order to derive effective bounds on the error of perturbed regularization we shall restrict to projection methods, below.
4. Projection methods

We shall further analyze the error for the case of projection methods given by \( B := Q_n AP_m \), for certain orthogonal projections \( Q := Q_n \) in \( Y \) and \( P := P_m \) in \( X \). In this case, \( \Theta(s_m^2) \) can directly be estimated as stated in

**Proposition 1** \[ \| A^* A - PA^* QAP \| \leq 2a \| A(I - P) \| + \| (I - Q)A \|^2. \]

For later use we add the following bound.

\[ \| PA^* AP - PA^* QAP \| \leq \| (I - Q)A \|^2. \]

We continue estimating the terms arising in the error decomposition as derived in § 3.3 in particular bounding the term \( (I - Q) \). First, it can trivially be estimated as

\[ \| B - QA : X_\varphi \to Y \| \leq \| A(I - P) : X_\varphi \to Y \|. \]

Again, for later use we add some estimate for for \( \| I - P : X_\varphi \to X \| \).

**Proposition 2** Let \( \varphi, \varphi(0) = 0 \), be any increasing index function. Then

\[ \| I - P : X_\varphi \to X \| \leq \begin{cases} \varphi(\| A(I - P) \|^2), & \text{if } \varphi^2 \text{ is concave}, \\ C_\varphi \| A(I - P) \|, & \text{if } X_\varphi \hookrightarrow X_{\sqrt{\ell}}, \end{cases} \]

where \( C_\varphi := \| X_\varphi \hookrightarrow X_{\sqrt{\ell}} \| \) is the norm of canonical embedding operator from \( X_\varphi \) to \( X_{\sqrt{\ell}} \). Consequently,

\[ \| A(I - P) : X_\varphi \to Y \| \leq \begin{cases} \| A(I - P) \| \varphi(\| A(I - P) \|^2), & \text{if } \varphi^2 \text{ is concave}, \\ C_\varphi \| A(I - P) \|^2, & \text{if } X_\varphi \hookrightarrow X_{\sqrt{\ell}}. \end{cases} \]

**Proof:** Applying Corollary 3 from Appendix A with \( T = T^* := I - P \), allows to prove (10). Since obviously,

\[ \| A(I - P) : X_\varphi \to Y \| \leq \| A(I - P) : X \to Y \| \| I - P : X_\varphi \to X \|. \]

the second statement follows from the first one. \( \square \)

**Remark 4** It can easily be seen, that

\[ \inf \{ \| A(I - P) : X_\varphi \to Y \|, \text{ rank}(P) < m \} = \Theta(s_m^2), \]

where \( \Theta(t) := \sqrt{t} \varphi(t) \), and \( s_m \) is \( m \)-th singular number of \( A \). Therefore, if the projection \( P \) is best possible, its cardinality \( m \) should be chosen, such that \( s_m^2 \approx \alpha = \Theta^{-1}(\delta) \). However, if discretization is involved, then the best estimate is provided by Proposition 4.

If \( P \) corresponds to a spectral cut-off, then the bound in (12) coincides with (12) only for index functions with concave square. Otherwise there is a saturation, and over-discretization is necessary to keep (11), hence (7), at the level \( \delta \).

We continue bounding the terms as indicated in § 3.3 using the notation \( B \) and \( QAP \), interchangeably. For \( \psi \in \mathcal{F}(d) \) the expression (8) can further be estimated as

\[ \| P\psi(A^* A) - \psi(B^* B) \| \]

\[ \leq \| P\psi(A^* A) - P\psi(A^* A)P \| + \| P\psi(A^* A)P - \psi(PA^* AP) \| \]

\[ + \| \psi(PA^* AP) - \psi(B^* B) \| \]

\[ \leq \| I - P : X_\psi \to X \| + d\psi(\| PA^* AP - B^* B \|) \]

\[ + \| P\psi(A^* A)P - \psi(PA^* AP) \|. \]
The first term on the right was estimated in Proposition 2. Also, the norm inside \( \psi \) was bounded in Proposition 1, such that we are left with bounding \( \| P\psi(A^*A)P - \psi(PA^*AP) \| \), properly.

To this end we need to further restrict the classes \( \mathcal{F}(d) \) and \( \mathcal{F}_L(d) \), involving operator monotone (concave) functions, we are thus led to the following classes.

\[
\tilde{\mathcal{F}}(d) := \{ \varphi, \varphi : (0, a) \to \mathbb{R}_+ \text{ operator monotone}, \varphi \in \mathcal{F}(d) \} \quad (14)
\]
\[
\tilde{\mathcal{F}}_L(d) := \{ \varphi \vartheta, \varphi : (0, a) \to \mathbb{R}_+, \varphi \in \tilde{\mathcal{F}}(d), \text{ and } \vartheta \text{ Lipschitz} \} \quad (15)
\]

The following result allows to bound the remaining term on the right in (13).

**Proposition 3** For any projection \( P \) and any function \( \psi \) from \( \tilde{\mathcal{F}}(d) \) we have

\[
\| P\psi(A^*A)P - \psi(PA^*AP) \| \leq d_1 \psi(\| A(I-P) \|^2). \quad (16)
\]

The proof will be given in Appendix B.

### 4.1. Error bounds

The estimates provided in Propositions 2 and 3 have different areas of validity. Proposition 2 needs to distinguish, whether \( \varphi^2 \) is concave or not. Proposition 3 required to restrict to functions arising from operator concave ones. Also, the structure of the error was different for functions from \( \mathcal{F}(d) \) and \( \mathcal{F}_L(d) \). Therefore, the obtained bounds provide different error behavior for four different classes of index functions, as these are made precise as follows.

\[
\mathcal{F}_0(d) := \{ \varphi \in \tilde{\mathcal{F}}(d), \varphi^2 \text{ is concave} \},
\]
\[
\mathcal{F}_{1/2}(d) := \{ \varphi \in \tilde{\mathcal{F}}(d), \varphi \leq c\sqrt{t} \text{ for some } c > 0 \},
\]
\[
\mathcal{F}_1(d) := \{ \varphi \in \tilde{\mathcal{F}}(d), \varphi = \vartheta \psi, \psi \in \mathcal{F}_0(d), \text{ and } \vartheta \text{ Lipschitz} \},
\]
and finally

\[
\mathcal{F}_{3/2}(d) := \{ \varphi \in \tilde{\mathcal{F}}(d), \varphi = \vartheta \psi, \psi \in \mathcal{F}_{1/2}(d), \text{ and } \vartheta \text{ Lipschitz} \}.
\]

It is now clear, how to bound the error for each of the classes \( \mathcal{F}_{p/2}(d), p = 0, 1, 2, 3. \)

To shorten notation we shall abbreviate the perturbations from the operator \( A \), induced by projections \( P \) from the right and \( Q \) from the left by

\[
\rho := \| A(I-P) \| \quad \text{and} \quad \eta := \| (I-Q)A \|,
\]
respectively. For given \( \varphi \in \tilde{\mathcal{F}}(d) \) let us introduce the following functions \( b_\varphi, v_\varphi : \rho \to \mathbb{R}_+ \) as

\[
b_\varphi(\rho) := \begin{cases} 
\varphi(\rho^2), & \varphi \in \mathcal{F}_0(d), \\
C_\varphi \rho, & \varphi \in \mathcal{F}_{1/2}(d),
\end{cases}
\]
and

\[
v_\varphi(\rho) := \begin{cases} 
\rho \varphi(\rho^2), & \varphi \in \mathcal{F}_0(d), \\
C_\varphi \rho^2, & \varphi \in \mathcal{F}_{p/2}(d), \quad p = 1, 2, 3.
\end{cases}
\]
Summarizing all estimates as provided in Propositions 1-3, we are now ready to state the main result on the structure of the error.

**Theorem 2** Let $B := QAP$ be a projection method with accuracies $\rho$ and $\eta$ from above. Suppose, that $g_\alpha$ is chosen, such that the qualification of $\varphi$ is covered. Then the following bound on the discretization error holds true.

(i) For index functions from the classes $\mathcal{F}_0(d)$ or $\mathcal{F}_{1/2}(d)$ we have

$$e(A_{\varphi}, g_\alpha(B^* B^*) B^*, \delta) \leq c \left( \varphi(\alpha) + \varphi(\eta^2) + \varphi(\rho^2) \right) + \gamma b_\varphi(\rho) + \frac{\gamma_s}{\sqrt{\alpha}} (\delta + v_\varphi(\rho)).$$

(ii) For index functions $\varphi$ from $\mathcal{F}_1(d)$ or $\mathcal{F}_{3/2}(d)$, which are represented as $\varphi = \vartheta \psi$ we can bound

$$e(A_{\varphi}, g_\alpha(B^* B^*) B^*, \delta) \leq c \left( \varphi(\alpha) + \vartheta(\alpha) \psi(\eta^2) + \vartheta(\alpha) b_\psi(\rho) \right) + \gamma (\eta^2 + 2a\rho)$$

$$+ \frac{\gamma_s}{\sqrt{\alpha}} (\delta + v_\varphi(\rho)).$$

**Remark 5** Suppose, the source condition $A_{\varphi}$ is given with a known function $\varphi$. Then, in order to maintain the best possible accuracy, it is sufficient to choose $\alpha := \Theta^{-1}(\delta)$ and $\rho \approx \eta \approx \delta$, since necessarily $\delta^2 = o(\alpha)$ as $\delta \to 0$.

For index functions $\varphi(t) := t^p$ this (non-adaptive) strategy was proposed in [9, Remark 1], since this allows to maintain the best possible accuracy regardless of the actual value of $p$.

### 4.2. Adaptive strategy

The above error estimates can be used to propose adaptive strategies in choosing the cardinality of data, thus the level $\eta$ as well as the required discretization level controlled by $\rho$. Our strategy is the analog to the one proposed in [7]. The idea is to choose along with $\alpha$ the levels $\rho$ and $\eta$ maximal, such that the error estimates from Theorem 1 are not spoiled. It is a routine matter to check, that this is achieved by choosing

$$\rho^2 := \rho \alpha^2(\alpha, \delta) := \min \left\{ \alpha, \frac{\delta^2}{\alpha} \right\},$$

and

$$\eta^2 := \eta \alpha^2(\alpha, \delta) := \min \left\{ \alpha, \frac{\delta}{\sqrt{\alpha}} \right\}.$$  \hspace{1cm} (17)

The typical behavior of these parameters as functions of $\alpha$ are shown in Figure 1.

Having determined the present values of the parameters $\rho, \eta$, we determine the projections $P$ and $Q$, respectively. The way this is done, depends on the approximative powers of these projections and is known in most cases. This choice results in a respective $B := B(\rho, \eta) := PAQ$. This adaptive strategy, including the adaptive choice of $\rho$ and $\eta$ is outlined schematically as shown in Figure 2.

**Remark 6** In this routine, each value of $\alpha$ corresponds to some (hidden) index function $\varphi$. Thus, as $\alpha$ increases, we (implicitly) travel through the classes
Figure 1. Choice of $\eta$ and $\rho$ as a function of $\alpha$. $\delta$ is chosen as $\delta = 0.1$.

**ADAPT**($q, C, \delta$)

$$
\alpha := \delta^2; \\
\eta := \eta(\alpha, \delta); \\
\rho := \rho(\alpha, \delta); \\
B := B(\rho, \eta); \\
x_{1,\delta} := g_\alpha(B^*B)B^*y_\delta;
$$

Do

$$
x_{0,\delta} := x_{1,\delta}; \\
\alpha := q \times \alpha; \quad / * \text{geom. progression} */ \\
\eta := \eta(\alpha, \delta); \\
\rho := \rho(\alpha, \delta); \\
B := B(\rho, \eta); \\
x_{1,\delta} := g_\alpha(B^*B)B^*y_\delta; \quad / * \text{regularize} */
$$

While ($\|x_{1,\delta} - x_{0,\delta}\| \leq C\delta \sqrt{\frac{q}{\alpha}} \| \alpha \leq \|A^*A\|$);

Return $x_{0,\delta}$;

Figure 2. The adaptive strategy
\(\mathcal{F}_0(d), \mathcal{F}_{1/2}(d), \mathcal{F}_1(d), \mathcal{F}_{3/2}(d)\) in the way, that to each value of \(\alpha\) corresponds a "true" source condition in one of the respective classes, for which \(c\varphi(\alpha) = \delta / \sqrt{\alpha}\). The adaptive strategy is chosen, such that the best possible accuracy is maintained throughout.

We summarize this discussion in the following corollary, and indicate how under the adaptive strategy the bounds from Theorem 2 can be rendered more precisely.

**Corollary 1** There are the constants \(C_1\) and \(C_{a,\gamma,\varphi}\), such that under the adaptive strategy from (17) and (18) the following bound holds true.

\[
e(A_{\varphi}, g_{\alpha}(B^*B)B^*, \delta) \leq C_1\varphi(\alpha) + C_{a,\gamma,\varphi} \frac{\delta}{\sqrt{\alpha}}.
\]

(19)

The constant \(C_1\) depends only on internal parameters such as \(d, d_1\), while \(C_{a,\gamma,\varphi} = 2a\gamma + (C_{\varphi} + 1)(\gamma + \gamma_\ast)\). By the choice of the regularization, \(C_{a,\gamma,\varphi} \leq C_{a,\gamma,\rho}\). The latter parameters are completely under our disposal and do not further depend on \(\varphi\). We even assume \(C_{a,\gamma,\rho} \geq 1\), for simplicity.

In contrast, the constant \(C_1\) depends on \(\varphi\), which is assumed to be unknown.

The bound (19) can be made uniform for index functions \(\varphi\), which are covered by the chosen regularization with qualification \(\rho\). It will provide the optimal order of accuracy, uniformly over the following class of index functions. Given \(0 < D < \infty\), let \(\mathcal{F}_D\) denote the class of all index function from \(\bigcup_{p=0}^3 \mathcal{F}_{p/2}(d)\), which additionally satisfy \(\varphi(qt) \leq D\varphi(t), \ t > 0\). This strategy is controlled by two parameters: \(q > 1\), describing the geometric progression from one value of \(\alpha\) to the next; and a valid bound \(C := 4C_{a,\gamma,\rho}\), which is given, once the qualification of \(g_{\alpha}\) is fixed.

**Theorem 3** On the class \(\mathcal{F}_D\) the adaptive strategy with initial parameters \((q, 4C_{a,\gamma,\rho}, \delta)\) provides the following error bound

\[
e(A_{\varphi}, g_{\alpha}(B^*B)B^*, \delta) \leq C_qD\sqrt{q}\varphi(\Theta^{-1}(\delta/C_1)), \ 0 < \delta \leq a,
\]

(20)

where \(C_q := 2C_{a,\gamma,\rho}C_1(1 + \frac{2\sqrt{q}}{\sqrt{q} - 1})\).

**Proof:** [Sketch] The same arguments as in the proof of Theorem 3 in [7] apply. But by Corollary 1 we start with the bound

\[
e(A_{\varphi}, g_{\alpha}(B^*B)B^*, \delta) \leq C_{a,\gamma,\rho} \left( C_1\varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}} \right).
\]

Inspecting of the proof of Proposition 4 from [7], using the unknown index function \(C_1\varphi\) instead of \(\varphi\), allows to conclude, that (20) is valid with constant \(C_q\) as stated. \(\square\)

5. Concluding discussion

Here we will discuss some features of the adaptive strategy as outlined above in some detail.
5.1. Complexity issues

The choices of \( \rho \) and \( \eta \) correspond to a necessary amount of discretization, expressed in terms of \( m \) and \( n \), and it is customary to assume, that both \( m \) and \( n \) are decreasing functions.

At any level \( \alpha \) the regularization \( g_\alpha(B^*B)B^*y_\delta \) requires to deal with a \( m \times n \)-matrix, where \( m = m(\rho) \) and \( n = n(\eta) \) are determined by the values of \( \rho \) and \( \eta \), respectively. Thus \( mn + n \) may be understood as the information complexity of the algorithm, but this is of the order \( mn \), which we shall study now. The computational cost is an (increasing) function of \( m \cdot n \).

First, recall that the choice of \( \rho \approx \eta \approx \delta \) is always sufficient, in order not to spoil the error bound. This finds its correspondence in \( \rho \geq \delta \) and \( \eta \approx \delta \), throughout. Therefore the quotient \( m(\rho)n(\eta)/m(\delta)n(\delta) \leq 1 \) and choosing \( \rho \) and \( \eta \) large is desirable.

For the choice according to (17), (18), and denoting briefly \( a \wedge b := \min \{a, b\} \), the gain is

\[
\frac{m(\sqrt{\alpha} \wedge \delta/\sqrt{\alpha})n(\sqrt{\alpha} \wedge \sqrt{\delta}/\sqrt{\alpha})}{m(\delta)n(\delta)}.
\]

In most cases, in particular for \( m(t) \approx t^{-r} \) for some \( r > 0 \), this quotient tends to 0, as \( \delta \to 0 \), since \( \alpha \) and \( \delta/\sqrt{\alpha} \) should tend to 0, as \( \delta \to 0 \). This is even more the case for the arguments of \( n \).

Secondly, if we have any a priori information on the minimal smoothness of \( \varphi \), e.g. \( X_\varphi \hookrightarrow X_{t\mu} \) for some \( \mu > 0 \), then this results in a choice of the initial value of \( \alpha_0 \) in the adaptive strategy, different from \( \delta^2 \). In particular, if we knew \( X_\varphi \hookrightarrow X_t \), then \( \alpha_0 := \delta^{2/3} \) is reasonable. This case is particularly important, since then we start with the coarsest discretization level and refine in the course of adaptation as long as necessary.

Previous studies, connecting discretization and/or adaptive choice of the regularization parameter, may be classified by two different approaches.

5.2. Discretization first

Here discretization is based on the noise level \( \delta \) and the regularization parameter \( \alpha \) is then chosen by some principle. We mention [9, 12] and more recently [8]. This approach causes problems. Discretization independently of \( \alpha \) may be too coarse to yield the optimal order of accuracy or unnecessarily fine, since depending on \( \delta \), only. Also, several a posteriori principles to choose \( \alpha \) are known not to cover the whole range of smoothness, where the chosen regularization may provide the optimal order of accuracy, as e.g. Morozov’s in connection with Tikhonov regularization; we refer to [13, Chapt. 3.3] for details, and also the discussion in [9, Thm. 3.3].

5.3. Discretization based on \( \alpha \)

In a second approach, given \( \alpha \), discretization is done depending on \( \alpha \) and/or \( \delta \). The classical paper in this direction is [2]. Below, let \( B \) denote a discretization of the original
A. In [2] Thm. 2] the condition \( \|A^*A - B^*B\| \asymp \alpha^2 \) is shown to allow the optimal rate of convergence for source conditions given by \( \varphi(t) = t \), thus belonging to \( \mathcal{F}_1(d) \) above. In case of projection methods this corresponds to \( \rho \asymp \eta^2 \asymp \alpha^2 \), which is finer than our choice, thus less efficient. In a recent study [10], this is improved to \( \|A^*A - B^*B\| \asymp \alpha \), still being inferior to the choice proposed in this paper.

In [5], the authors study Tikhonov regularization and propose \( \|A - B\|^2 \asymp \alpha \delta \), which in the context of projection methods translates to \( \rho \asymp \eta \asymp \sqrt{\alpha \delta} \). Again, it can easily be checked, that on the class of source conditions, where Tikhonov regularization provides the best possible order of accuracy, the choice of \( \rho \) and \( \eta \) according to (17), (18) is superior. This may be due to the fact, that discretization based on distances of \( A \) from \( B \) is not appropriate. Instead, as is suggested by (17), (18), discretization from the left and right may differ.

5.4. A posteriori choice of the regularization parameter

As mentioned above, the a posteriori choice of the regularization parameter by several of the known principles may not yield the optimal order of accuracy for given source conditions. The strategy as proposed in this paper automatically provides the optimal order of accuracy for all index functions covered. For the precise statement we refer to Theorem 3. But notice, that the levels \( \rho \) and \( \eta \) have to be chosen as functions of \( \alpha \) (and \( \delta \)).

Summarizing, the choice of \( \rho \) and \( \eta \) as described in (17) and (18), respectively, is the largest known one, which yields the best order of accuracy. Moreover, the adaptation strategy we propose, see the schematic view in Figure 2, yields the optimal order of accuracy over the whole range of source conditions, covered by the chosen regularization.

Appendix A. Interpolation in variable Hilbert scales

The analysis, in particular for estimating (7), requires some basic results on interpolation, which we shall present in this section. We first recall, that given any index function \( \varphi \) on the spectrum of \( A^*A \), we assign a Hilbert space \( X_\varphi \), having \( A_\varphi \) as its unit ball.

Precisely, \( A^*A \) admits a (monotonic) Schmidt representation for an orthonormal system \( u_1, u_1, \ldots \), given by

\[
A^*Ax = \sum_{j=1}^{\infty} s_j \langle x, u_j \rangle u_j, \quad x \in X.
\]

Then \( X_\varphi \) is the completion of finite expansions \( x = \sum_{j=1}^{n} \langle x, u_j \rangle u_j \) with respect to the scalar product

\[
\langle x, y \rangle := \sum_{j=1}^{\infty} \frac{\langle x, u_j \rangle \langle y, u_j \rangle}{\varphi^2(s_j)}.
\]
Corollary 2 Under the assumptions from Theorem 4 it holds

Theorem 4 has an operator analog, which can be stated as thus

\[ \|x\|_{\theta/\varphi} \leq \psi^{-1}\left( \frac{\|x\|_{\theta/\psi}}{\|x\|_{\theta}} \right), \quad x \in X_{\max\{\varphi,\psi\}^{-1}\theta} \setminus 0. \]  \quad (A.1)

Proof: Let \( x \in X_{\max\{\varphi,\psi\}^{-1}\theta} \setminus 0 \). Without loss of generality we assume \( \|x\|_{\theta} = 1 \). Furthermore, let \( \xi_1, \xi_2, \ldots \) denote the Fourier coefficients of \( x \) with respect to the Eigenbasis of \( A^*A \). Then \( \|x\|^{2}_{\theta/\psi} := \sum_{j=1}^{\infty} \psi^2(s_j)\xi_j^2/\theta^2(s_j) \) is a convex combination. The concavity assumption implies

\[
\varphi^{-1}\left( \frac{\|x\|_{\theta/\varphi}}{\|x\|_{\theta}} \right) \leq \psi^{-1}\left( \frac{\|x\|_{\theta/\psi}}{\|x\|_{\theta}} \right), \quad \varphi^2(\psi^2)^{-1}\left( \psi^2(s_j)\xi_j^2/\theta^2(s_j) \right) \geq \sum_{j=1}^{\infty} \varphi^{-1}\left( \frac{\|x\|_{\theta/\psi}}{\|x\|_{\theta}} \right) = \sum_{j=1}^{\infty} \varphi(\psi^2)^{-1}\left( \psi^2(s_j)\xi_j^2/\theta^2(s_j) \right) = \|x\|^2_{\theta/\varphi}.
\]

Taking \((\varphi^2)^{-1}\) and square roots allows to complete the proof. \( \square \)

Remark 7 Under the assumptions of Theorem 4 necessarily

\[ X_{\theta/\psi} \hookrightarrow X_{\theta/\varphi} \hookrightarrow X_{\theta}, \]

which explains the term interpolation.

Example 1 Let \( r \leq s \leq t \) and \( \vartheta \) be such that \( s = \vartheta r + (1 - \vartheta)t \). Then, with \( \theta(\alpha) := \alpha^t \), \( \varphi(\alpha) := \alpha^{t-s} \) and \( \psi(\alpha) := \alpha^{t-r} \) we see, that \( \varphi^2(\psi^2)^{-1} := \alpha^{(t-s)/(t-r)} \) is concave.

Applying \( (A.1) \) we arrive at

\[ \frac{\|x\|_{s}}{\|x\|_{t}} \leq \left( \frac{\|x\|_{r}}{\|x\|_{t}} \right)^{\vartheta}, \]

thus \( \|x\|_{s} \leq \|x\|_{t}^{\vartheta} \|x\|_{t}^{1-\vartheta} \), where \( \|x\|_{s} \) denotes \( \|x\|_{t*}, \ldots \)

Theorem 4 has an operator analog, which can be stated as

Corollary 2 Under the assumptions from Theorem 4 it holds

\[ \varphi^{-1}\left( \frac{\|T: X_{\theta} \rightarrow Y\|}{\|T: X_{\vartheta} \rightarrow Y\|} \right) \leq \psi^{-1}\left( \frac{\|T: X_{\theta} \rightarrow Y\|}{\|T: X_{\theta} \rightarrow Y\|} \right), \]  \quad (A.2)

Proof: To adapt to the setup of Theorem 4 we rename \( \theta := 1/\vartheta \). Let \( y_0 \in Y \) be chosen such that \( \|T^*y_0\|_{\theta/\varphi} = \|T^* : Y \rightarrow X_{\theta/\varphi}\|, \|y_0\| = 1 \). The interpolation inequality \( (A.1) \) yields

\[ \frac{\|T^*y_0\|_{\theta/\varphi}^2}{\|T^*y_0\|_{\theta}^2} \leq \varphi^2\left( (\psi^2)^{-1}\left( \frac{\|T^*y_0\|_{\theta/\psi}^2}{\|T^*y_0\|_{\theta}^2} \right) \right), \]
which implies
\[ \|T^*y_0\|_{\tilde{\theta}/\varphi} \leq \|T^*y_0\|_{\tilde{\theta}}^2 \varphi^2 \left( \left( \frac{1}{\psi^2} \right) \left( \frac{\|T^*y_0\|_{\tilde{\theta}/\varphi}^2}{\|T^*y_0\|_{\tilde{\theta}}^2} \right) \right). \]

Now, since for any concave function \( f: [0, a] \to \mathbb{R}_+ \), \( f(0) = 0 \), we have \( sf(x/s) \leq tf(x/t) \), for all \( 0 \leq s \leq t \), whenever, \( x/t \leq a \) and since \( \|T^*y_0\|_{\tilde{\theta}} \leq \|T^*: Y \to X_{\theta}\| \), the above argument implies
\[ \frac{\|T^*: Y \to X_{\theta/\varphi}\|^2}{\|T^*: Y \to X_{\theta}\|^2} \leq \varphi^2 \left( \left( \frac{1}{\psi^2} \right) \left( \frac{\|T^*y_0\|_{\tilde{\theta}/\varphi}^2}{\|T^*y_0\|_{\tilde{\theta}}^2} \right) \right), \]
from which the proof easily can be completed. \( \square \)

As an important application we mention the following result, cf. [1, Problem IX.8.14]

**Proposition 4 (Generalization of the Peierls-Bogolyubov-Inequality)** Let \( Z \) be any Hilbert space and \( T: X \to Z \) be any operator bounded in norm by 1. If \( \varphi \) is any index function such that \( \varphi^2 \) is concave, then
\[ \|T: X_\varphi \to Z\| \leq \varphi(\|AT^*: Z \to Y\|^2). \]  

**Proof**: If \( \varphi \) is any index function such that \( \varphi^2 \) is concave, then necessarily \( X_{\sqrt{t}} \hookrightarrow X_\varphi \). Applying (A.2) with \( \theta = 1 \) and \( \psi(t) = \sqrt{t} \) we arrive at
\[ \|T: X_\varphi \to Z\| \leq \varphi(\|T: X_{\sqrt{t}} \to Z\|)^2. \]

Since, for a certain \( z \) with \( \|z\| = 1 \),
\[ \|T: X_{\sqrt{t}} \to Z\|^2 = \|T^*: Z \to X_{1/\sqrt{t}}\|^2 = \langle \sqrt{A^*AT^*}z, \sqrt{A^*AT^*}z \rangle = \langle A^*AT^*z, T^*z \rangle = \langle AT^*z, AT^*z \rangle \leq \|AT^*: Z \to Y\|^2, \]
we obtain (A.3). \( \square \)

The “minimal” index functions \( \varphi \), for which their squares are concave, are multiples of \( t \to \sqrt{t} \). Thus, if \( X_\varphi \hookrightarrow X_{\sqrt{t}} \), then \( \|T: X_\varphi \to Z\| \leq C\|T: X_{\sqrt{t}} \to Z\| \). Summarizing, we arrive at

**Corollary 3** Let \( \varphi, \varphi(0) = 0 \), be any increasing index function. Then, for \( \|T\| \leq 1 \), it holds with \( C_\varphi := \|X_\varphi \hookrightarrow X_{\sqrt{t}}\|
\[ \|T: X_\varphi \to Z\| \leq \begin{cases} \varphi(\|AT^*: Z \to Y\|^2), & \text{if } \varphi^2 \text{ is concave,} \\ C_\varphi \|AT^*: Z \to Y\|, & \text{if } X_\varphi \hookrightarrow X_{\sqrt{t}}. \end{cases} \]

**Appendix B. Operator concave functions**

As already mentioned, the terms (8), thus (16), involve the respective index function \( \psi \) at different operator arguments. To bound such norms adequately, the notion of operator concave functions is important. First analysis in this direction was undertaken by the present authors in [6]. This will be used here.

We start with the following variation of Theorem 1 in [6]. Recall, that a non-negative operator monotone function on \( (0, \infty) \) is operator concave there.
Theorem 5 Let \( \varphi \) operator monotone on \((0, \infty)\) with \( \varphi(0) = 0 \). Then for any projection \( P \) it holds

\[
\| P \varphi(A^*A)P - \varphi(PA^*AP) \| \leq 2 \varphi(\|A(I-P)\|^2). \tag{B.1}
\]

Proof: Plainly, (B.1) is true for linear functions \( t \to \beta t \). It is well known, that every function \( \varphi \), operator monotone on \((0, \infty)\), with \( \varphi(0) = 0 \), admits a representation \( \varphi(t) := \beta t + \int_0^\infty \frac{t}{\lambda + t} \mu(d\lambda) \), for arbitrary (non-negative) measures \( \mu \), for which \( \int_0^\infty (1 + \lambda)^{-1} \mu(d\lambda) < \infty \), see [1], Eq. V.53 and also [6], Sec. 1. Therefore, it is enough to show (B.1) for functions \( f_\lambda(t) := t(\lambda + t)^{-1} \), the building stones of operator monotone on \((0, \infty)\) functions. Since for every \( P \) we have \( f_\lambda(PA^*AP) \geq Pf_\lambda(A^*A)P \), see e.g., [1], Thm. V.2.3, we can bound

\[
\| Pf_\lambda(A^*A)P - f_\lambda(PA^*AP) \| \leq \| f_\lambda(PA^*AP) \| \leq 1,
\]

which is useful for small \( \lambda \). Moreover, from the representation

\[
Pf_\lambda(A^*A)P - f_\lambda(PA^*AP) = -PA^*(\lambda I + AA^*)^{-1}A(I-P)A^*(\lambda I + APA^*)^{-1}AP,
\]

we deduce

\[
\| Pf_\lambda(A^*A)P - f_\lambda(PA^*AP) \| \leq \| A^*(\lambda I + AA^*)^{-1}\|\| A(I-P)A^* \| \| (\lambda I + APA^*)^{-1}AP \|
\leq \| A(I-P)A^* \| \leq \| A(I-P) \|^2,
\]

useful for \( \lambda \geq \| A(I-P) \|^2 \). Altogether we can bound

\[
\| Pf_\lambda(A^*A)P - f_\lambda(PA^*AP) \| \leq \min \left\{ 1, \frac{\| A(I-P) \|^2}{\lambda} \right\}
\leq 2 \frac{\| A(I-P) \|^2}{\lambda + \| A(I-P) \|^2} = 2f_\lambda(\| A(I-P) \|^2),
\]

thus completing the proof.

As in [6, Theorem 2] we are going to extend the range of applicability to more functions, those as defined in \( \mathcal{F}(d) \). This was stated in Proposition 3 in §4, which we are going to prove now.

Proof: [Proof of Proposition 3] Recall from [6], that every function \( \varphi \) from \( \mathcal{F}(d) \) can be written as a sum \( \varphi = \varphi_0 + \varphi_1 \), of a function \( \varphi_0 \), which obeys the assumptions of Theorem 5 and a function \( \varphi_1(t) := \int_a^\infty \frac{t}{(\lambda - 0)^2} \mu(d\lambda) \), for some finite measure \( \mu \) obeying \( \int (\lambda^2 + 1)^{-1} \mu(d\lambda) < \infty \). This, and since \( S\varphi(S^*S) = \varphi(SS^*)S \), for operators \( S \), allows to bound

\[
\| P\varphi_1(A^*A)P - \varphi_1(PA^*AP) \|
\leq \int_a^\infty \| PA^*(\lambda I - AA^*)^{-1}AP - PA^*(\lambda I - APA^*)^{-1}AP \| \frac{\mu(d\lambda)}{\lambda}
\leq \int_a^\infty \|PA^*(\lambda I - AA^*)^{-1}(APA^* - AA^*)(\lambda I - APA^*)^{-1}AP \| \frac{\mu(d\lambda)}{\lambda}
\leq \| A(I-P) \|^2 \int_a^\infty (\lambda - \| A \|^2)^{-2} \mu(d\lambda) \leq C_1 \| A(I-P) \|^2.
\]
But, the operator concave part $\varphi_0$ is concave and increasing, which implies $t \leq C_2 \varphi_0(t)$, for some constant $C_2$. This allows to complete the proof. \hfill \Box

**Remark 8** Actually, it is very easy to establish

$$
\| P(A^* A)^n P - (P A^* A P)^n \| \leq C \| A(I - P) \|^2,
$$

for monomials $t \rightarrow t^n$, $n \geq 1$, and some constant $C = C(\| A \|)$. Therefore, the estimate extends to arbitrary polynomials $\sum_{j=1}^{n} c_j t_j$, for which $\sum_{j=1}^{n} |c_j|$ is uniformly bounded, independent of $n$. This approach would also cover the above proof for the function $\varphi_1$.

It would be of interest to know, how far estimates of this type can be extended.

**References**


