A new regularization method for linear exponentially ill-posed problems.

Walter Cedric SIMO TAO LEE

September 23, 2019

Abstract

This paper provides a new regularization method which is particularly suitable for linear exponentially ill-posed problems. Under logarithmic source conditions (which have a natural interpretation in terms of Sobolev spaces in the aforementioned context), concepts of qualifications as well as order optimal rates of convergence are presented. Optimality results under general source conditions expressed in terms of index functions are also studied.

Finally, numerical experiments on three test problems strongly attest the superiority of the new method compared to the well known Tikhonov method in instances of exponentially ill-posed problems.

Keywords: Ill-posed problems, Regularization, logarithmic source conditions, qualifications, order-optimal rates.

1 Introduction

In this paper, we are interested in the solution to the equation

$$Tx = y$$

(1)

where $T : X \to Y$ is a linear bounded operator between two infinite dimensional Hilbert spaces $X$ and $Y$ with non-closed range. The data $y$ belongs to the range of $T$ and we assume that we only have approximated data $y^\delta$ satisfying

$$||y^\delta - y|| \leq \delta.$$  

(2)

In such a setting, Equation (1) is ill-posed in the sense that the operator $T^\dagger$ which maps $y$ to the solution $x^\dagger$ of (1) is not continuous. Consequently a little perturbation on the data $y$ may induce an arbitrarily large error in the solution $x^\dagger$. Instances of such ill-posed inverse problems are encountered in several fields in applied sciences among which: signal and image processing, computer tomography, immunology, satellite gradiometry, heat conduction problems, inverse scattering problems, statistics and econometrics to name just a few (see, e.g. [7, 9, 13, 14, 25]). As a result of the ill-posedness of Equation (1), a regularization method needs to be applied in order to recover from the noisy data $y^\delta$ a stable approximation $x^\delta$ of the solution $x^\dagger$. A regularization method can be regarded as a
family of continuous operators $R_\alpha : Y \to X$ such that there exists a function $\Lambda : \mathbb{R}_+ \times Y \to \mathbb{R}_+$ satisfying the following: for every $y = Tx^\dagger \in \text{Ran } T$ and $y^\delta \in Y$ satisfying (2),

$$R_{\Lambda(\delta, y^\delta)} y^\delta \to x^\dagger \quad \text{as } \delta \downarrow 0. \quad (3)$$

Some examples of regularization methods are Tikhonov, Landweber, spectral cut-off, asymptotic regularization, approximate inverse and mollification (see, e.g. [1, 6, 7, 14, 17, 18]). As a matter of fact, we would like to get estimates on the error committed while approximating $x^\dagger$ by $x^\delta = R_{\Lambda(\delta, y^\delta)} y^\delta$.

It is well known that for arbitrary $x^\dagger \in X$, the convergence of $x^\delta$ towards $x^\dagger$ is arbitrary slow (see, e.g. [7, 29]). But still, by allowing smoothness of the solution $x^\dagger$, convergence rates could be established. Standard smoothness conditions known as Hölder type source condition take the form

$$x^\dagger \in X_\mu(\rho) = \{(T^*T)^\mu w, \quad w \in X \quad \text{s.t. } ||w|| \leq \rho\}, \quad (4)$$

where $\mu$ and $\rho$ are two positive constants. However such source conditions have shown their limitations as they are too restrictive in many problems and do not yield a natural interpretation. For this reason, general source conditions have been introduced in the following form:

$$x^\dagger \in X_\varphi(\rho) = \{\varphi(T^*T)w, \quad w \in X \quad \text{s.t. } ||w|| \leq \rho\}, \quad (5)$$

where $\rho$ is a positive constant and $\varphi : \sigma(T^*T) \to \mathbb{R}_+$ is an index function, i.e. a non-negative monotonically increasing continuous function defined on the spectrum $\sigma(T^*T)$ of $T^*T$ and satisfying $\varphi(\lambda) \to 0$ as $\lambda \downarrow 0$. An interesting discussion on these source conditions can be found in [23] where the author explores how general source conditions of the form (5) are. Once the solution $x^\dagger$ satisfies a smoothness condition i.e. $x^\dagger$ belongs to a proper subspace $M$ of $X$, it is possible to derive convergence rates and the next challenge is about optimality. More precisely, for a regularization method $R : Y \to X$, we are interested in the worst case error:

$$\Delta(\delta, R, M) := \sup \left\{||Ry^\delta - x^\dagger||, \quad x^\dagger \in M, \quad y^\delta \in Y, \quad \text{s.t. } ||y^\delta - Tx^\dagger|| \leq \delta \right\}, \quad (6)$$

and we would like a regularization which minimizes this worst case error. In this respect, a regularization method $\bar{R} : Y \to X$ is said to be optimal if it achieves the minimum worst case error over all regularization methods, i.e. if

$$\Delta(\delta, \bar{R}, M) = \Delta(\delta, M) := \inf_{\bar{R}} \Delta(\delta, R, M).$$

Similarly, a regularization is said to be order optimal if it achieves the minimum worst case error up to a constant greater than one, i.e. if

$$\Delta(\delta, \bar{R}, M) \leq C \Delta(\delta, M).$$

for some constant $C > 1$. When the subset $M$ is convex and balanced, it is shown in [24] that

$$\omega(\delta, M) \leq \Delta(\delta, M) \leq 2 \omega(\delta, M), \quad (7)$$

where $\omega(\delta, M)$ is the modulus of continuity of the operator $T$ over $M$ i.e.

$$\omega(\delta, M) = \sup \left\{||x||, \quad x \in M, \quad \text{s.t. } ||Tx|| \leq \delta \right\}. \quad (8)$$
In other words, we get the following:

\[ \Delta(\delta, X_\varphi(\rho)) = \mathcal{O}(\omega(\delta, X_\varphi(\rho))). \]  

(9)

Recall that, under mild assumptions on the index function \( \varphi \), the supremum defining
the modulus of continuity is achieved and a simple expression of \( \omega(\delta, X_\varphi(\rho)) \) in term of
function \( \varphi \) is available (see, e.g. [13, 22, 31]). Let us remind that a relevant notion in the
study of optimality of a regularization method is qualification. In fact, the qualification
of a regularization measures the capability of the method to take into account smoothness
assumptions on the solution \( x^\dagger \), i.e. the higher the qualification, the more the method is
able to provide best rates for very smooth solutions.

Besides optimality, converse results and saturation results are also important aspects of
regularization algorithms (see, [7, 21, 27, 28]). For converse results, we are interested in the
following: given a particular convergence rate of \( \|x^\delta - x^\dagger\| \) towards 0, which smoothness
condition does the solution \( x^\dagger \) needs to satisfy? Saturation results are about the maximal
smoothness on the solution \( x^\dagger \) for which a regularization method can still deliver the best
rates of convergence. Finally, another significant aspect of regularization is the selection of the
regularization parameter i.e. finding a function \( \Lambda(\delta, y^\delta) \) which guarantees convergence
and possibly order-optimality.

Coming back to (5), notice that a very interesting subclass of general source conditions
are logarithmic source conditions expressed as:

\[ x^\dagger \in X_{f_p}(\rho) = \left\{ (-\ln(T^*T))^{-p}w, \quad w \in X \quad \text{s.t.} \quad \|w\| \leq \rho \right\}, \]  

(10)

where \( p \) and \( \rho \) are positive constants. Such smoothness conditions have clear interpreta-
tions in term of Sobolev spaces in exponentially ill-posed problems (see, [13, 31]). The
latter class includes several problems of great importance such as backward heat equation,
sideways heat equation, inverse problem in satellite gradiometry, control problem in heat
equation, inverse scattering problems and many others (see, [13]). Because of the impor-
tance of exponentially ill-posed problems, it is desirable to design regularization methods
particularly suitable for this class of problems. It is precisely the aim of this paper to
provide such a regularization scheme.

In the next section, we define the new regularization method using both the variational
formulation and the definition in terms of the so called generator function \( g_\alpha \). A brief
comparison with the Tikhonov method is done. Moreover basic estimates on the generator
function \( g_\alpha \) and on the residual function are also carried out.

Section 3 is devoted to optimality of the new method. Here we recall well known
optimality results under general source conditions of the form (5) (see, [12, 13, 22, 26, 31]).
For the specific case of logarithmic source conditions, qualification of the method is given
and order optimality is shown. Next we study optimality under general source conditions.

In Section 4, we present a comparative analysis of the new method with the Tikhonov
method and the spectral cut-off.

Section 5 is about numerical illustrations. In this section, in order to confirm our
prediction of superiority of the new method compared to Tikhonov and spectral cut-off
in instance of exponentially ill-posed problems, we numerically compare the efficiency of
the three methods on three test problems coming from literature: A problem of image
reconstruction taken from [30], a Fredholm integral equation of the first kind found in [2]
and an inverse heat equation problem.

Finally in Section 6, for a fully applicability of the new method, we exhibit heuristic
selection rules which fit with the new regularization technique.
2 The new regularization method

For the sake of simplicity, we assume henceforth that the operator $T$ is injective.

Let us consider the general variational formulation of a regularization method

$$x_\alpha = \arg\min_{x \in X} F(Tx, y) + P(x, \alpha)$$

where $F(Tx, y)$ is the fit term, $P(x, \alpha)$ is the penalty term and $\alpha > 0$ is the regularization parameter. We recall that the fit term aims at fitting the model, the penalty term aims at introducing stability in the initial model $Tx = y$ and the regularization parameter $\alpha$ controls the level of regularization.

In most cases, the fit term $F(Tx, y)$ is nothing but

$$F(Tx, y) = ||Tx - y||^2_Y.$$  

and the penalty term depends on the regularization method. For instance, for Tikhonov regularization, $P(x, \alpha)$ is given by

$$P(x, \alpha) = \alpha ||x||^2_X.$$  

This penalization can sometimes compromises the quality of the resulting approximated solution $x_\alpha$. Indeed, let $X = L^2(\mathbb{R}^n)$, then by Parseval identity, we see that

$$P(x, \alpha) = \alpha ||\hat{x}||^2_{L^2(\mathbb{R}^n)}$$

where $\hat{x}$ is the Fourier transform of $x$. Equation (14) implies that the stability is introduced by penalizing all frequency components irrespective of the magnitude of frequencies. Yet, it is well known that the instability of the initial problem comes from high frequency components on the contrary to low frequency components.

Let us introduce the following penalty term where the regularization parameter $\alpha$ is no more defined as a weight but as an exponent:

$$P(x, \alpha) = \left\| \left[ I - (T^*T)^{\sqrt{\alpha}} \right] x \right\|^2_X.$$  

In (15), $(T^*T)^{\sqrt{\alpha}}$ is defined via the spectral family $(E_\lambda)_\lambda$ associated to the self-adjoint operator $T^*T$ i.e.

$$(T^*T)^{\sqrt{\alpha}}x = \int_{\lambda=0}^{||T^*T||} \lambda^{\sqrt{\alpha}} dE_\lambda x.$$  

We keep the fit term defined in (12) and then the variational formulation of our new regularization method is given by

$$x_\alpha = \arg\min_{x \in X} ||Tx - y||^2_Y + \left\| \left[ I - (T^*T)^{\sqrt{\alpha}} \right] x \right\|^2_X.$$  

From the first order optimality condition, we get that $x_\alpha$ is the solution to the linear equation :

$$\left[T^*T + \left( I - (T^*T)^{\sqrt{\alpha}} \right)^2 \right] x = T^*y,$$

that is

$$x_\alpha = \left[T^*T + \left( I - (T^*T)^{\sqrt{\alpha}} \right)^2 \right]^{-1} T^*y.$$
From (17), we see that the new method can also be defined via the so called generator function $g_\alpha$, i.e.
\[ x_\alpha = g_\alpha(T^*T)T^*y, \tag{18} \]
with the function $g_\alpha$ defined by
\[ g_\alpha(\lambda) = \frac{1}{\lambda + (1 - \lambda \sqrt{\alpha})^2}, \quad \lambda \in (0, ||T^*T||]. \tag{19} \]
Let us also define the residual function $r_\alpha$ as follows
\[ r_\alpha(\lambda) := 1 - \lambda g_\alpha(\lambda) = \frac{(1 - \lambda \sqrt{\alpha})^2}{\lambda + (1 - \lambda \sqrt{\alpha})^2}, \quad \lambda \in (0, ||T^*T||]. \tag{20} \]

The functions $g_\alpha$ and $r_\alpha$ defined in (19) and (20) are important since they will be repeatedly used in the convergence analysis of the regularization method. In fact, the regularization error $x^\dagger - x_\alpha$ and the propagated error $x_\alpha - x_\alpha^\delta$ are expressed via the functions $r_\alpha$ and $g_\alpha$ as follows:
\[ x^\dagger - x_\alpha = r_\alpha(T^*T)x^\dagger, \quad x_\alpha - x_\alpha^\delta = g_\alpha(T^*T)T^*(y - y^\delta). \]

Finally, notice that the function $g_\alpha$ defined in (19) indeed satisfies the basic requirements for defining a regularization method i.e.

a) $g_\alpha$ is continuous,

b) $\forall \alpha > 0, \sup_{\lambda \in (0, ||T^*T||]} \lambda g_\alpha(\lambda) \leq 1 < \infty$,

c) $\lim_{\alpha \downarrow 0} g_\alpha(\lambda) = 1/\lambda$.

From b) and c), we deduce the convergence of the new regularization method by application of [7, Theorem 4.1]. Before going to optimality results, let us state some basic estimates (proven in the appendix) about the functions $g_\alpha$ and $r_\alpha$.

**Proposition 1.** Let the function $g_\alpha$ be defined by (19). Then for all $a < 1$ and $\alpha < 1$,
\[ \sup_{\lambda \in (0, a]} \sqrt{\lambda} g_\alpha(\lambda) = O\left(\frac{1}{\sqrt{\alpha}}\right). \tag{21} \]

**Lemma 1.** For all $\alpha$ and $\lambda$ satisfying $0 < \alpha \leq \lambda < 1$, the following estimates hold for the function $r_\alpha$ defined in (20):
\[ r_\alpha(\lambda) \leq \frac{9}{4} \left( \frac{\alpha |\ln(\lambda)|^2}{\lambda + \alpha |\ln(\lambda)|^2} \right). \tag{22} \]

### 3 Optimality results

Before studying the optimality of the method presented in Section 2, we first need to recall general optimality results under source condition of the form (5). For doing so, let us specify assumptions on the function $\varphi$ which defines the source set $X_\varphi(\rho)$. Hereafter, we set a positive number $a$ such that the operator norm of $T^*T$ is less than $a$ i.e. $||T^*T||_{\mathcal{L}(X)} \leq a$.

**Assumption 1.** The function $\varphi : (0, a] \rightarrow \mathbb{R}_+$ is continuous, monotonically increasing and satisfies:

(i) $\lim_{\lambda \downarrow 0} \varphi(\lambda) = 0$,
(ii) the function $\phi : (0, \varphi^2(a)] \to (0, a\varphi^2(a)]$ defined by
\[ \phi(\lambda) = \lambda^{(\varphi^2)}^{-1}(\lambda) \] (23)
is convex.

Under Assumption 1 on the function $\varphi$, the following result from [31] holds and we can then defines optimality under source condition (5).

**Theorem 1.** Let $X_\varphi(\rho)$ be as in (5) and let Assumption 1 be fulfilled. Let the function $\phi$ be defined by (23). Then
\[ \omega(\delta, X_\varphi(\rho)) \leq \rho \sqrt{\phi^{-1}(\frac{\delta^2}{\rho^2})}. \] (24)
Moreover, if $\delta^2/\rho^2 \in \sigma(T^*T\varphi^2(T^*T))$, then equality holds in (24).

A similar result to this theorem can be found in [13, Section 2], and [22, Section 3].

**Remark 1.** In the reference [22], the results corresponding to Theorem 1 are given in term of the function $\Theta : (0, a] \to (0, a\varphi(a)]$ defined by:
\[ \Theta(\lambda) = \sqrt{\lambda}\varphi(\lambda). \] (25)
Then, by simple computations, we can find that
\[ \rho \sqrt{\phi^{-1}(\frac{\delta^2}{\rho^2})} = \rho \varphi(\Theta^{-1}(\delta/\rho)). \] (26)
In such a case, the convexity of the function $\phi$ defined in (23) is equivalent to the convexity of the function $\chi(\lambda) = \Theta((\varphi^2)^{-1}(\lambda))$ and the condition $\delta^2/\rho^2 \in \sigma(T^*T\varphi^2(T^*T))$ which allows to get the equality in (24) is equivalent to $\delta/\rho \in \sigma(\Theta(T^*T))$.

From Theorem 1 and Remark 1, we can deduce that under the source condition (5) and Assumption 1, the best possible worst case error is $\rho \varphi(\Theta^{-1}(\delta/\rho))$ whence the following definition.

**Definition 1** (Optimality under general source conditions). Let Assumption 1 be satisfied and consider the source condition $x^1 \in X_\varphi(\rho)$. A regularization method $R(\delta) : Y \to X$ is said to be:

- **optimal** if $\Delta(\delta, R(\delta), X_\varphi(\rho)) \leq \rho \varphi(\Theta^{-1}(\delta/\rho))$;
- **order optimal** if $\Delta(\delta, R(\delta), X_\varphi(\rho)) \leq C \rho \varphi(\Theta^{-1}(\delta/\rho))$ for some constant $C \geq 1$;
- **quasi-order optimal** if for all $\epsilon > 0$, $\Delta(\delta, R(\delta), X_\varphi(\rho)) = O(f_\epsilon(\delta))$ where the function $f_\epsilon : \mathbb{R}_+ \to \mathbb{R}_+$ converges to $\varphi(\Theta^{-1}(\delta/\rho))$ as $\epsilon$ decreases to 0 i.e. for all $\delta > 0$, $f_\epsilon(\delta) \to \varphi(\Theta^{-1}(\delta/\rho))$ as $\epsilon$ decreases to 0.

Having defined the optimality under general source conditions, let us now consider the particular case of logarithmic source conditions. For logarithmic source conditions, the function $\varphi$ equals the function $f_p : [0, a] \to \mathbb{R}_+$ defined by:
\[ f_p(\lambda) = (-\ln(\lambda))^{-p}. \] (27)
Next it is easy to see that the only point to check in Assumption 1 is the convexity of the function $\phi$ defined in (23). Precisely, for the index function $f_p$, this function is $\phi_p : (0, \ln(1/a)^{-2p}] \to (0, a \ln(1/a)^{-2p}]$ defined by
\[ \phi_p(\lambda) = \lambda \exp(-\lambda^{-1/2p}) \]
which was proven to be convex on the interval $[0, 1]$ in [20]. So in order to fulfill Assumption 1 and avoid the singularity of the function $f_p$ at $\lambda = 1$, we assume that $a \leq \exp(-1) < 1$, i.e. $||T^*T||_{\mathcal{X}(X)} \leq \exp(-1)$. Notice that this is not actually a restriction, since Equation (1) can always be rescaled in order to meet this criterion.

Due to (26) it suffices to compute $\sqrt{\phi_p^{-1}(\delta^2/\rho^2)}$ in order to define the optimality in logarithmic source conditions. Thanks again to [20], we have that

$$\sqrt{\phi_p^{-1}(s)} = f_p(s)(1 + o(1)) \text{ as } s \to 0. \quad (28)$$

Hence, we deduce the following definition of optimality in case of logarithmic source condition.

**Definition 2** (Optimality under logarithmic source condition). Consider logarithmic source condition (10), on defining $f_p$ as in (27), a regularization method $R(\delta) : Y \to X$ is said to be:

- **optimal** if $\Delta(\delta, R(\delta), X_{f_p}(\rho)) \leq f_p(\delta^2/\rho^2)(1 + o(1))$ as $\delta \to 0$,
- **order optimal** if $\Delta(\delta, R(\delta), X_{f_p}(\rho)) \leq C f_p(\delta^2/\rho^2)(1 + o(1))$ as $\delta \to 0$.

where the function $f_p$ is defined in (27).

### 3.1 Optimality under logarithmic source conditions

Having given all the necessary definitions, let us now study the optimality of the method proposed in Section 2.

**Proposition 2.** The regularization $g_\alpha$ defined by (19) has qualification $f_p$. That is:

$$\sup_{0 < \lambda \leq a} |r_\alpha(\lambda)| f_p(\lambda) = \mathcal{O}(f_p(\alpha)). \quad (29)$$

The proof the Proposition 2 heavily relies on the following lemma which is proven in the appendix.

**Lemma 2.** Let $p$ and $\alpha$ be two positive numbers, let $a \in (0, 1)$ and let $\Psi_{p, \alpha} : (0, a] \to \mathbb{R}_+$ be the function defined by

$$\Psi_{p, \alpha}(\lambda) = \frac{|\ln(\lambda)|^{2-p}}{\lambda + \alpha |\ln(\lambda)|^2}. \quad (30)$$

Then, the followings hold:

(i) The function $\Psi_{p, \alpha}$ is well defined and derivable on $(0, 1)$, and its derivative is given by

$$\Psi'_{p, \alpha}(\lambda) = \frac{x^{-1}|\ln(\lambda)|^{1-p}}{(\lambda + \alpha |\ln(\lambda)|^2)^2} h(\lambda), \quad (31)$$

where

$$h(\lambda) = \alpha p |\ln(\lambda)|^2 - \lambda (2 - p + |\ln(\lambda)|). \quad (32)$$

(ii) If $p < 2$, there exists at least one $\lambda(\alpha, p)$ which cancels the function $h$. Moreover for every such $\lambda(\alpha, p)$, the following holds

$$\lambda(\alpha, p) \simeq \alpha |\ln(\alpha)|, \quad (33)$$

that is, there exists two constants $c_1$ and $c_2$ depending on $p$ only such that

$$c_1 \alpha |\ln(\alpha)| \leq \lambda(\alpha, p) \leq c_2 \alpha |\ln(\alpha)|.$$

Moreover, this result still holds if $p \geq 2$, $\lambda < c \leq \exp(2 - p)$ and $\alpha$ is small.
(iii) The supremum of the function $\Psi_{p,\alpha}$ on $(0, a]$ satisfies
\[
\sup_{0<\lambda \leq \alpha} \Psi_{p,\alpha}(\lambda) = \mathcal{O}\left(\alpha^{-1}|\ln(\alpha)|^{-p}\right). \tag{34}
\]

Having stated the above lemma, the proof of Proposition 2 easily follows:

**Proof.** If $\lambda \leq \alpha$ then the monotonicity of the function $f_p$ and the fact that the residual function $r_\alpha$ is bounded by 1 on $(0, a]$ yields (29). If $\lambda \geq \alpha$ then from Lemma 2, we deduce that
\[
\sup_{0<\lambda \leq \alpha} \alpha|\ln(\lambda)|^2 f_p(\lambda) = \mathcal{O}\left(f_p(\alpha)\right)
\]
which together with Lemma 1 yields (29).

From Proposition 2, we deduce the following optimality result.

**Theorem 2.** Let $p > 0$, $x^\dagger \in X_{f_p}(\rho)$, and $y^\delta \in Y$ satisfy (2) with $y = T x^\dagger$. Assume that $||T^*T||_{\mathcal{L}(X)} \leq \exp(-1)$ and let $x(\delta) = g_\alpha(\delta)(T^*T)^{-1}y^\delta$ with the function $g_\alpha$ being defined by (19) and let $\alpha(\delta) = \Theta_p^{-1}(\delta)$ with $\Theta_p$ defined by
\[
\Theta_p(\lambda) = \sqrt{\lambda}(\ln(1/\lambda))^{-p}. \tag{35}
\]
Then the order optimal estimate
\[
||x^\dagger - x(\delta)|| = \mathcal{O}\left(f_p(\delta)\right) \tag{36}
\]
holds as $\delta \to 0$. Thus the regularization $g_\alpha$ defined by (19) is order optimal under logarithmic source conditions.

**Proof.** As usual, we start with the following splitting
\[
||x^\dagger - x^\delta\alpha|| \leq ||x^\dagger - x_\alpha|| + ||x_\alpha - x^\delta\alpha||. \tag{37}
\]
Using that $x^\dagger - x_\alpha = r_\alpha(T^*T)x^\dagger$, $x_\alpha - x^\delta = g_\alpha(T^*T)(y - y^\delta)$ together with the source condition $x^\dagger \in X_{f_p}(\rho)$, we deduce that:
\[
||x^\dagger - x_\alpha|| \leq C_1 \sup_{\lambda \in (0, a]} r_\alpha(\lambda) f_p(\lambda) \tag{38}
\]
and
\[
||x_\alpha - x^\delta\alpha|| \leq \delta C_2 \sup_{\lambda \in (0, a]} \sqrt{\lambda} g_\alpha(\lambda). \tag{39}
\]
By applying the propositions 1 and 2 to (38), (39) and using (37), we get that
\[
||x^\dagger - x^\delta\alpha|| \leq C'_1 f_p(\alpha) + C'_2 \frac{\delta}{\sqrt{\alpha}}, \tag{40}
\]
where $C'_1$ and $C'_2$ are constants independent of $\alpha$ and $\lambda$. Hence, by taking $\alpha := \Theta_p^{-1}(\delta)$, we deduce (36).

A converse result
Theorem 2 establishes that the logarithmic source condition (10) is sufficient to imply the rate \( f_p(\delta) \) in (36). Now we are going to prove that the logarithmic source condition (10) is not only sufficient but also almost necessary. The following result based on [13, Theorem 8] establishes a converse result in the noise free case for the new regularization method.

**Theorem 3.** Let \( x_\alpha = g_\alpha(T^*T)Ty \) with \( y = Tx^\dagger \) and let the function \( g_\alpha \) be defined in (19). Then the estimates

\[
\|x^\dagger - x_\alpha\| = \mathcal{O}(f_p(\alpha)) \tag{41}
\]

imply that \( x^\dagger \in X_{g_\alpha}(\rho) \) for some \( \rho > 0 \) for all \( 0 < q < p \).

The proof consists in checking that the function \( g_\alpha \) defined in (19) satisfies all the conditions stated in Theorem 8 of [13]. More precisely, we just need to check that there exists a constant \( C_g > 0 \) such that

\[
\sup_{\lambda \in (0,||T^*T||]} g_\alpha(\lambda) \leq \frac{C_g}{\alpha}.
\]

But, from (52), we see that the latter condition is obviously fulfilled.

### 3.2 Optimality under general source conditions

Let us state the following quasi-optimal result under general source conditions.

**Theorem 4.** Let \( p > 0 \), \( x^\dagger \in X_\varphi(\rho) \), where \( \varphi \) is a concave index function satisfying Assumption 1 and \( y^\theta \in Y \) satisfying \( \|y - y^\theta\| \leq \delta \) with \( y = Tx^\dagger \). Assume that \( ||T^*T||_{\varphi(X)} \leq a \leq \exp(-1) \) and let \( x(\delta) = g_\alpha(\delta)(T^*T)y^\delta \) with the function \( g_\alpha \) defined in (19). For small positive \( \epsilon \), let \( \alpha(\delta) = \Theta^{-1}_\epsilon(\delta) \) with the function \( \Theta_p \) defined by \( \Theta_p(\lambda) = \lambda^{-\epsilon}\Theta(\lambda) \) with the function \( \Theta \) defined in (25).

Then the estimate

\[
\|x^\dagger - x(\delta)\| = \mathcal{O}((\Theta^{-1}_\epsilon(\delta))^{-\epsilon}\varphi(\Theta^{-1}_\epsilon(\delta)))
\]

holds as \( \delta \to 0 \). Moreover, as \( \epsilon \downarrow 0 \), \( (\Theta^{-1}_\epsilon(\delta))^{-\epsilon}\varphi(\Theta^{-1}_\epsilon(\delta)) \to \varphi(\Theta^{-1}(\delta)) \). Thus the regularization method defined via the function \( g_\alpha \) given in (19) is quasi-order optimal under general source conditions.

**Proof.** We study two cases: \( \alpha \geq \lambda \) and \( \alpha < \lambda \). In the first case, \( \sup_{(0,\exp^{-1})} r_\alpha(\lambda)\varphi(\lambda) \leq \varphi(\alpha) \) by monotonicity of the function \( \varphi \) and the order-optimality follows trivially. Let us study the main case when \( \alpha < \lambda \). From Lemma 1, we get, for \( \lambda \in (0, a] \),

\[
\begin{align*}
r_\alpha(\lambda)\varphi(\lambda) & \leq \frac{9}{4} \left| \ln(\lambda) \right|^2 \frac{\alpha}{\lambda + \alpha|\ln(\lambda)|^2} \varphi(\lambda) \\
& \leq \frac{9}{4} \left| \ln(\lambda) \right|^2 \frac{\alpha}{\lambda + \alpha|\ln(\lambda)|^2} \varphi(\lambda) \\
& \leq \frac{9}{4} \alpha^{-\epsilon} (\alpha^{\epsilon/2}|\ln(\lambda)|)^2 \frac{\alpha}{\lambda + \alpha|\ln(\lambda)|^2} \varphi(\lambda) \\
& \leq \frac{9}{4} \frac{\alpha \lambda}{\alpha \lambda + \alpha|\ln(\lambda)|^2} \frac{\varphi(\lambda)}{\lambda} \\
& \leq \frac{9}{4} \frac{\alpha \lambda}{\alpha \lambda + \alpha|\ln(\lambda)|^2} \frac{\varphi(\alpha)}{\alpha} \quad \text{by concavity of } \varphi \\
& \leq C_\epsilon \alpha^{-\epsilon} \varphi(\alpha).
\end{align*}
\]
Hence \(\sup_{[0,a]} r_{\alpha}(\lambda) \varphi(\lambda) \leq C_\epsilon \alpha^{-\epsilon} \varphi(\alpha)\). From (38) and (39), and (21) we get
\[
||x^\dagger - x^\delta_\alpha|| \leq C_\epsilon \alpha^{-\epsilon} \varphi(\alpha) + \frac{\delta}{\sqrt{\alpha}}.
\]
By taking \(\alpha(\delta) = \Theta^{-1}_\epsilon(\delta)\) with \(\Theta_\epsilon(\lambda) = \lambda^{1/2-\epsilon} \varphi(\lambda)\), we get
\[
||x^\dagger - x(\delta)|| = O((\Theta^{-1}_\epsilon(\delta))^{-\epsilon} \varphi(\Theta^{-1}_\epsilon(\delta)))\,.
\]
Now, it remains to show that \((\Theta^{-1}_\epsilon(\delta))^{-\epsilon} \varphi(\Theta^{-1}_\epsilon(\delta)))\) converges to the optimal rate \(\varphi(\Theta^{-1}_\epsilon(\delta))\) as \(\epsilon\) goes to 0.

Let \(\alpha_* = \Theta^{-1}_\epsilon(\delta)\) and \(\alpha_\epsilon = \Theta^{-1}_\epsilon(\delta)\), let us show that \(\alpha_\epsilon\) converges to \(\alpha_*\) as \(\epsilon\) goes to 0. Since \(\alpha_\epsilon \in (0,a]\), the sequence \((\alpha_\epsilon)_\epsilon\) is bounded and thus it admits a converging subsequence. Let \((\alpha_{\epsilon_n})_n\) a converging subsequence of \((\alpha_\epsilon)_\epsilon\), and let \(\alpha\) be its limit. Let us show that \(\alpha = \alpha_*\).

Since \(\alpha_{\epsilon_n} \to \alpha\) and \(\Theta\) is continuous, \(\Theta(\alpha_{\epsilon_n}) \to \Theta(\alpha)\). But \(\Theta(\alpha_{\epsilon_n}) = \alpha_{\epsilon_n}^\epsilon \Theta(\alpha_*)\) since \(\delta = \Theta(\alpha_*)\) and \(\delta = \Theta(\alpha_\epsilon)\) for all small positive \(\epsilon\). So we get
\[
\alpha_{\epsilon_n}^\epsilon \Theta(\alpha_*) \to \Theta(\alpha)\text{ i.e. } \alpha_{\epsilon_n}^\epsilon \to \frac{\Theta(\alpha)}{\Theta(\alpha_*)}.
\]
By the convergence of the sequence \((\alpha_{\epsilon_n})_n\), we get that \(\alpha_{\epsilon_n}^\epsilon = \exp(\epsilon_n \ln(\alpha_{\epsilon_n}))\) converges to 1, (42) proves that \(\Theta(\alpha) = \Theta(\alpha_*)\) and by bijectivity of the function \(\Theta\), we deduce that \(\alpha = \alpha_*\). Since the sequence \((\epsilon_n)_n\) was arbitrarily chosen, we deduce that the whole sequence \((\alpha_\epsilon)_\epsilon\) converges to \(\alpha_*\) as \(\epsilon \downarrow 0\). Thus we deduce that \(\alpha_\epsilon^{-\epsilon} \to 1\) and \(\varphi(\alpha_\epsilon) \to \varphi(\alpha_*)\) which implies that
\[
(\Theta^{-1}_\epsilon(\delta))^{-\epsilon} \varphi(\Theta^{-1}_\epsilon(\delta)) \to \varphi(\Theta^{-1}_\epsilon(\delta)).
\]
\[\square\]

For Holder type source conditions, Theorem 4 reduces to the following theorem.

**Theorem 5.** Consider the setting of Theorem 4 with the function \(\varphi(t) = t^\mu\) i.e. \(x^\dagger \in \text{Ran}(T^*T)^\mu\), then there exists an a priori selection rule \(\alpha(\delta)\) such that the following holds:
\[
||x^\dagger - x^\delta_\alpha|| = \begin{cases} 
O\left(\frac{2\sigma}{\delta^{\frac{\sigma+1}{\mu}}}\right) & \forall \sigma < \mu, \text{ if } \mu \leq 1 \\
O\left(\frac{2}{\delta^2}\right) & \text{, if } \mu > 1.
\end{cases}
\]
(43)

**Remark 2.** By defining a variant of the new regularization method where the approximated solution \(x^\delta_\alpha\) is defined as the solution of the optimization problem
\[
x_\alpha = \arg\min_{x \in X} ||(T^*T)^{\alpha} y^\delta - Tx||_Y^2 + \frac{||I - (T^*T)^{\alpha}||}{2} x||_X^2,
\]
we can prove order optimal rate under Holder type source condition but with a lower qualification index \(\mu_0 = 1/2\). This variant is motivated by the mollification regularization method, where a target object defined as a smooth version of \(x^\dagger\) is fixed prior to the regularization (see e.g. [1, 6]). In this respect, the target object here is given as \((T^*T)^{\alpha} x^\dagger\). This choice is legitimated by the smoothness property of the operator \(T\) and the fact that as \(\alpha\) goes to 0, this target object converges to the solution \(x^\dagger\). The study of this variant and the corresponding optimality results is beyond the scope of this paper.
4 A framework for comparison

Let us first put the three regularization methods to compare (new method, Tikhonov \[32\] and spectral cut-off) into a common framework.

4.1 Unifying framework

A natural frame in which the three regularization methods under considerations (Tikhonov, new method and spectral cut-off) fit is the frame of general regularization method defined via a generator function. Roughly speaking, each regularization method is defined via a so-called generator function $g_{\alpha}^{\text{reg}}(\lambda)$ which converges pointwise to $1/\lambda$ as $\alpha$ goes to 0 and the approximation solution $x_{\alpha,\text{reg}}^\delta$ is defined by:

$$x_{\alpha,\text{reg}}^\delta = g_{\alpha}^{\text{reg}}(T^*T)T^*y^\delta.$$ (44)

In this respect, the functions $g_{\alpha}^{\text{reg}}(\lambda)$ associated to Tikhonov, spectral cut-off and the new method are defined as follows:

$$
g_{\alpha}^{\text{tik}}(\lambda) = \frac{1}{\lambda + \alpha}, \quad g_{\alpha}^{\text{norm}}(\lambda) = \frac{1}{\lambda + (1 - \lambda \sqrt{\alpha})^2}, \quad g_{\alpha}^{\text{sc}}(\lambda) = \frac{1}{\lambda}1(\lambda \geq \alpha),$$ (45)

where $\lambda \in (0, a]$ with $\|T^*T\|_{\mathcal{L}(X)} \leq a < 1$.

Let us now see how each regularization method proceeds in order to introduce boundedness in the operator $g_{\alpha}^{\text{reg}}(T^*T)$ which approximates the unbounded operator $(T^*T)^{-1}$.

4.2 Tikhonov regularization

In order to avoid the blow-up of the function $1/\lambda$ at $\lambda = 0$, the Tikhonov method proceeds by shifting its curve by $\alpha$ along the $x$-axis. This is equivalent to shifting all singular values by $\alpha$ in order to avoid their accumulation to 0 which is the source of ill-posedness. By doing so, this method treat all frequency component in the same way without distinguishing high frequency components from low frequency components. This fact is also visible from the variational formulation of Tikhonov regularization where the penalty term is $\alpha ||x||_{X}$. Indeed, by considering $X = L^2(\mathbb{R}^n)$ for instance, by using the Parseval identity, we see that the penalty term is equal to $\alpha ||\hat{x}||_{L^2(\mathbb{R}^n)}$, meaning that the weight $\alpha$ equally penalizes all frequency components irrespective of the magnitude of frequencies. This is actually a drawback as it is well known that the cause of instability are usually only high frequency components especially in exponentially ill-posed problems. Consequently, this may induce an unfavorable trade-off between stability and fidelity to the model.

4.3 Spectral cut-off

For spectral cut-off, the blow-up of the function $1/\lambda$ at $\lambda = 0$ is avoided by doing a multiplication by a window function vanishing near $\lambda = 0$. This method, on the contrary to Tikhonov, treats high frequency components and low frequency components in a different way. Indeed, low frequency components are kept unchanged and regularization is applied to high frequency components by a mere cut-off. However, this method also presents a drawback as it completely gets rid of high frequency components. Indeed, even though those frequencies induces instability, they also carry some information which should not completely left out. The truncation introduced by the method is too crude and violent in some sense. For instance, for exponentially ill-posed problem, this truncation introduced...
by spectral cut-off will be less damaging on the quality of the approximation while for mildly ill-posed problems, this truncation will be more damaging. A smooth transition (in term of regularization) from small singular values to other singular values would be more meaningful and desirable.

4.4 New method

For the new method, in order to avoid the blow-up of the function $1/\lambda$ at $\lambda = 0$, the term $(1 - \lambda \sqrt{\alpha})^2$ (which equals 1 at $\lambda = 0$) is added to $\lambda$ in the denominator. Notice that the regularization term $(1 - \lambda \sqrt{\alpha})^2$ depends on $\lambda$ and this term is decreasing as a function of $\lambda$. Hence, for small values of $\lambda$ (i.e. for small singular values), this term increases and saturates at 1 while for $\lambda$ approaching 1, this term goes to 0. This means that for the new regularization method, high frequency components are much more regularized compared to low frequency components which are less and less regularized as the singular values increase to 1. In this way, we expect the new method to achieve a better trade-off between stability and fidelity to the model. Moreover, for exponentially ill-posed problems, the ill-posedness is accentuated due to the magnitude of singular values, the instability introduced by high frequency components are more pronounced and we expect the new regularization method to yield better approximations of the solution $x^\dagger$ compared to Tikhonov and spectral cut-off.

A comparative plot of the generator functions $g^\alpha_{\text{reg}}$ associated to each regularization method is given in Figure 1.

![Comparison of generator functions](image)

Figure 1: comparison generator function $g^\alpha_{\text{reg}}$ to function $\lambda \to 1/\lambda$ for the three regularization methods (reg = tik,sc,nrm).

**Remark 3.** The generator function $g^\alpha_{\text{nrm}}$ associated to the new regularization always exhibits a maximum close to $\lambda = 0$ and the function always equals 1 at $\lambda = 0$. Hence, the function $g^\alpha_{\text{nrm}}$ can be seen as a smooth version of the function $g^\alpha_{\text{sc}}$ associated to spectral cut-off which has a very crude transition at $\lambda = \alpha$.

5 Numerical illustration

The aim here is to compare the performance of our new regularization method (nrm) to the classical Tikhonov method (tik) and spectral cut-off (tsvd) for some (ill-posed) test problems. We consider three test problems. The first one is a problem of image reconstruction found in [30]. The second problem is a Fredholm integral equation of the first kind
taken from [2] and the last one is an inverse heat problem. For the discretization of these problems, we use the functions \texttt{shaw()}, \texttt{baart()} and \texttt{heat()} of the \texttt{matlab} regularization tool package (see [11]).

We consider a 4\% noise level, the noise level being defined here by the ratio of the noise to the exact data ($||\xi||/||y||$). We perform a Monte Carlo experiment of 3000 replications. In each replication, we compute the best relative error for each regularization method. Next we compute the minimum, maximum, average and standard deviation errors (denoted by $e_{\text{min}}$, $e_{\text{max}}$, $\bar{e}$, $\sigma(e)$ over the 3000 replications for each schemes (\texttt{nrm} and \texttt{tik} and \texttt{tsvd}). Figure 2 summarizes the results of the overall simulations.

In order to assess and compare the trade-off between stability and fidelity to the model for each regularization method, we plot the curve of the conditioning versus relative error. The conditioning here is the condition number of the reconstructed operator $g_{\alpha}^{\text{reg}}(T^*T)$ associated to the regularization method. For instance, using the invariance of conditioning by inversion, for the new method, the conditioning corresponds to the condition number of the operator $T^*T + \left[I - (T^*T)\sqrt{\alpha}\right]^2$ while for Tikhonov method, it corresponds to the condition number of $T^*T + \alpha I$. In this respect, for two regularization methods, the best one is the one whose curve is below the other one as it achieves the same relative errors with smaller conditioning. On Figure 3, for each test problem, we compare the curve of conditioning versus relative error of the new method and Tikhonov method.

Notice that the first two problems (\texttt{shaw} and \texttt{baart}) are midly ill-posed while the third problem (\texttt{heat}) is exponentially ill-posed.

|          | baart |         |          |         | heat |
|----------|-------|---------|----------|---------|------|
|          |       |         |          |         |      |
| $e_{\text{max}}$ | 0.3486 | 0.3632 | 0.3475 | 0.1902 | 0.2122 | 0.1898 | 0.3003 | 0.3171 | 0.2930 |
| $e_{\text{min}}$ | 0.0566 | 0.0527 | 0.1171 | 0.0486 | 0.0479 | 0.0577 | 0.1104 | 0.1162 | 0.1154 |
| $\bar{e}$    | 0.1736 | 0.1791 | 0.2044 | 0.1417 | 0.1440 | 0.1607 | 0.2048 | 0.2154 | 0.2123 |
| $\sigma(e)$  | 0.0453 | 0.0503 | 0.0445 | 0.0326 | 0.0353 | 0.0169 | 0.0301 | 0.0302 | 0.0301 |

Figure 2: Summary of the Monte Carlo experiment. On the upper figure, the average relative error for each method is represented by the vertical stick.

Comments:

- From Figure 2, we see that the new method always yields the smallest error among the three method. However, for the two midly ill-posed problems (\texttt{shaw} and \texttt{baart}), its errors are not significantly smaller than those of tikhonov on the contrary to the exponentially ill-posed problem (\texttt{heat}) where the new method produces smaller error than Tikhonov (about 5\% smaller). This confirms our prediction about the better performance of the new method in instance of exponentially ill-posed problems.
- For all three test problems, the new method performs better than spectral cut-off as could be expected. Moreover, the gap between the error is larger for the first two
Figure 3: Comparison of the trade-off between stability and accuracy of the new method (nrm) to Tikhonov (tik) for the three test problems: shaw, baart and heat.

test problems which are midly ill-posed. This also confirms the prediction about the poor performance of spectral cut-off for midly ill-posed problems.

- On the contrary to the two midly ill-posed problems (shaw and baart), spectral cut-off performs better than Tikhonov on the last test problem (heat), which is exponentially ill-posed. This emphasizes, especially in exponentially ill-posed problems, the drawback of Tikhonov method which regularizes all frequency component in the same way.

- From Figure 3, we can see that the new method achieves a better trade-off between stability and fidelity to the model compared to the Tikhonov method. Indeed, for the three test problems the curve associated to the new method lies below the one of Tikhonov. This means that given a stability level $\kappa$ (measured in term of conditioning), the new method provided a smaller error than Tikhonov. Conversely, for a given error level $\epsilon$, the new method provides a lower conditioning of the reconstructed operator compared to Tikhonov. This also validates the prediction stated earlier.

6 Parameter selection rules

In this section, we are interested in the choice of the regularization parameter $\alpha$ for the new regularization method. For practical purposes, we assume that we don’t know the smoothness conditions satisfied by the unknown solution $x^\dagger$. Consequently, we are left with two types of parameter choice rules: A-posteriori rules which use information on the noise level $\delta$ and heuristic rules which depend only on the noisy data $y^\delta$. An interesting converging a-posteriori parameter choice rule that could suit for the new method is the one proposed in [22] by Mathé and Pereverzev which is based on the Lepskii approach [16]. The first step of the method consists in discretizing the regularization parameter as follows:

$$\alpha_n = \alpha_0 q^n, \text{ with } \alpha_0 > \delta^2, q > 1, \text{ and } n = 1, 2, \ldots, N$$

where $N$ satisfies $\alpha_{N-1} \leq \|T^*T\| \leq \alpha_N$. Next the regularization parameter $\bar{\alpha}$ is chosen as

$$\bar{\alpha} := \max \left\{ \alpha_i, \ |x^\delta_{\alpha_i} - x^\delta_{\alpha_{i-1}}| \leq 4C_\gamma \gamma \sqrt{\alpha_{i-1}} \right\},$$

(46)

where $C_\gamma = \max\{\gamma_*, \bar{\gamma}\}$. Here $\gamma_*$ is the constant on the right hand side in (21) (which can be computed) and $\bar{\gamma}$ is the constant on the right hand side in the qualification inequality, e.g. for logarithmic source conditions, $\bar{\gamma}$ is the constant in the right hand side of (29).
However, for practical purposes, the unavailability of the constant $\bar{\gamma}$ for the new method and the fact that the noise level is hardly available or well estimated makes this rule difficult to use in our context. We turn to Heuristic (or data driven) selection rules. We recall that, due to Bakushinskii’s veto, such rules are not convergent. But still, heuristic rules may yield better approximations compared to sophisticated a-posteriori rules (see e.g. [10]) and this is not surprising as the Bakushinskii results is based on worst case scenario.

We applied four noise-free parameter choice rules to the new method on the three test problems defined in Section 5: the generalized cross validation (GCV), the discrete quasi-optimality rule (DQO) and two others heuristic rules (H1 and H2) described in [7, Section 4.5]. Roughly speaking, the parameter $\alpha$ chosen by each of those selection rules is as follows:

- The GCV rule consists in choosing $\hat{\alpha}$ as
  \[
  \hat{\alpha} = \arg\min_{\alpha} \frac{||Tx_\alpha^\delta - y^\delta||}{\text{tr}(r_\alpha(T^*T))},
  \]
  where $r_\alpha$ is the residual function associated to the regularization method under consideration. For the new method, $r_\alpha$ is defined in (20).

- The DQO method consists in discretizing the regularization parameter $\alpha$ as
  \[
  \alpha_n = \alpha_0 q^n, \quad \alpha_0 \in (0, ||T^*T||], \quad 0 < q < 1.
  \]
  Next, the parameter $\hat{\alpha}$ is chosen as
  \[
  \hat{\alpha} = \alpha_{\hat{n}} \quad \text{with} \quad \hat{n} = \arg\min_{n \in \mathbb{N}} ||\alpha_n \frac{\partial x_\alpha^\delta}{\partial n}|| = ||x_{\alpha_{n+1}}^\delta - x_{\alpha_n}^\delta||.
  \]
  (47)

Recall that this rule defined by (47) is actually one of the variants of the continuous quasi-optimality rule defined by
  \[
  \hat{\alpha} = \arg\min_{\alpha} ||\alpha \frac{\partial x_\alpha^\delta}{\partial \alpha}||.
  \]

- The third rule H1 taken in [7, Section 4.5] consists in choosing the parameter $\hat{\alpha}$ as
  \[
  \hat{\alpha} = \arg\min_{\alpha} \frac{1}{\alpha} ||Tx_\alpha^\delta - y^\delta||.
  \]
  (48)

- The fourth rule H2 which is a variant of the third rule H1 consists in choosing the parameter $\hat{\alpha}$ as
  \[
  \hat{\alpha} = \arg\min_{\alpha} \frac{1}{\alpha} ||T^*(Tx_\alpha^\delta - y^\delta)||.
  \]
  (49)

For a comprehensive discussion of the above heuristic rules and conditions under which convergence is established, see [33, 8, 19] for GCV, [15, 3, 4, 5] for Quasi-optimality and [7, Section 4.5] for the rules H1 and H2.

For assessing the performance of each selection rule, we perform a Monte Carlo experiment of 3000 replications. For each replication and each test problem (baart, shaw, heat), we compute the optimal regularization parameter $\alpha_{OPT}$, the one chosen by each selection rule ($\alpha_{GCV}, \alpha_{DQO}, \alpha_{H1}, \alpha_{H2}$). We also compute the corresponding relative errors:

\[
\frac{||x^\dagger - x_{\alpha_{OPT}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{GCV}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{DQO}}^\delta||}{||x^\dagger||}, \quad \frac{||x^\dagger - x_{\alpha_{H1}}^\delta||}{||x^\dagger||}, \quad \text{and} \quad \frac{||x^\dagger - x_{\alpha_{H2}}^\delta||}{||x^\dagger||}.
\]

In order to analyse the convergence behavior of the selection rules, we consider two noise levels $\delta$: 2% and 4%. The results are shown in the tables 1 and 2 and Figure 4.

From Tables 1, 2 and Figure 4, we can see the following:
Table 1: Summary of the Monte carlo experiment with the four heuristic rules GCV, DQO, H1 and H2 applied to the test problems shaw and baart.

| Test Problem | 4% nl | 2% nl |
|--------------|-------|-------|
|                | e_{max} | e_{min} | e | σ(e) | α |
| shaw          | OPT      | GCV     | DQO | H1  | H2  | OPT      | GCV     | DQO | H1  | H2  |
| 0.1892       | 0.1959   | 0.3496   | 0.5929 | 4.8475 | 0.3540 | 0.3768 |
| 0.1416       | 0.1654   | 0.1743   | 0.3663 | 0.2190 | 0.3455 | 0.3111 |
| 0.0328       | 0.0103   | 0.0473   | 0.6537 | 0.2012 | 0.0021 | 0.0075 |
| 1.42e-3      | 7.17e-4  | 4.2e-2   | 8.5e-4 | 2.17e-5 | 4.03e-2 | 3.8e-3 |
| heat         | OPT      | GCV     | DQO | H1  | H2  | OPT      | GCV     | DQO | H1  | H2  |
| 0.2948       | 0.3414   | 0.9267   | 0.9267 |
| 0.0970       | 0.1350   | 0.9247   |
| 0.0310       | 0.0063   | 0.0323   | 0.3613 | 0.0188 | 0.0010 | 0.0023 |
| 9.22e-7      | 8.5e-7   | 1.326    |

Table 2: Summary of the Monte carlo experiment with the four heuristic rules GCV, DQO, H1 and H2 applied to the new method for the test problem heat.

| Test Problem | 4% nl | 2% nl |
|--------------|-------|-------|
|                | e_{max} | e_{min} | e | σ(e) | α |
| heat          | OPT      | GCV     | DQO | H1  | H2  | OPT      | GCV     | DQO | H1  | H2  |
| 0.2948       | 0.3414   | 0.9267   |
| 0.0970       | 0.1350   |
| 0.0310       | 0.0063   |
| 9.22e-7      | 8.5e-7   |

Figure 4: Comparison of the relative $L^2$ error obtained by each selection rules (GCV, DQO, H1 and H2) for the two noise levels with the new method for the three tests problems shaw, baart and heat. On each plot, the x-axis corresponds to relative $L^2$ error and the vertical stick indicates the average relative error.

- For the exponentially ill-posed problem heat, from Table 2 and the last column of Figure 4, we can see that the errors produced by the DQO rule are very near the optimal error. Hence, the discrete quasi-optimality rule provides very good regularization parameter. Moreover the rule is very stable with respect to variations of the
error term in $y$. Next, the GCV rule exhibit an average $L^2$ relative error near the optimal one, however the GCV is not stable with respect to the noise in $y$ and this is shown by the spreading of dots along the x-axis or the corresponding large standard deviation $\sigma(e)$. Finally, the rules H1 and H2 always produce very large regularization parameters and diverge.

- For the middly ill-posed test problems shaw and baart, the rules H1 and H2 are very stable and produce good regularization parameters. The DQO rule yields average relative $L^2$ error close to the optimal one but is not stable except for the baart test problem with 2% noise level. The same applies to the GCV rule which is always unstable.

- For all the test problems, the DQO and GCV rules exhibit a convergence behavior. Indeed, for each test problem, the average chosen regularization parameter $\bar{\alpha}$ and the average relative error $\bar{e}$ corresponding to DQO and GCV always decrease as the noise error decreases from 4% to 2%. The same behavior holds for the rules H1 and H2 for the two middly ill-posed test problems shaw and baart. The only cases left are the rules H1 and H2 applied to the heat test problem where obviously, divergence occurs.

In summary, for the new regularization method, the discrete quasi-optimality rule is a good heuristic selection rule for exponentially ill-posed problems while the rules H1 or H2 might be preferable for middly ill-posed problems.

7 Conclusion

We presented a new regularization method which is particularly suitable for linear exponentially ill-posed problems. We study convergence analysis of the new method and we provided order optimal convergence rates under logarithmic source conditions which has a natural interpretation in term of Sobolev spaces for exponentially ill-posed problems. For a general source condition expressed via index functions, we only provided quasi-optimal rates. We saw that the new method performs better than the Tikhonov method and spectral cut-off for the considered exponentially ill-posed problem. For the two middly ill-posed problems considered, we saw that the new method actually gives results which are quite similar as those of the Tikhonov regularization. We finally applied some data-driven selection rules to the new method and the results suggest that the discrete quasi-optimality rule is a very efficient parameter choice rule for the new method in the context of exponentially ill-posed problem. In the context of middly ill-posed problems, the results of experiments suggest that the rules H1 and H2 described in Section 6 are preferable.

An interesting perspective would be the theoretical analysis of the discrete quasi-optimality rule (resp. rules H1 and H2) in the framework of exponentially (resp. middly) ill-posed problems in order to shed light on their good performances.

Acknowledgements: The author would like to thank Pierre Maréchal and Anne Vanhems for their helpful comments, readings and remarks.
8 Appendix

Proof of Proposition 1. Let us state the following standard inequality that we will use in the sequel:

\[ \forall t \geq 0, \quad \exp(-t) \leq \frac{1}{1 + t}. \]  

(50)

Using (50) applied with \( t = -\sqrt{\alpha \ln(\lambda)} \geq 0 \), we get

\[ 1 - \exp(\sqrt{\alpha \ln(\lambda)}) \geq 1 - \frac{1}{1 - \sqrt{\alpha \ln(\lambda)}} = -\frac{\sqrt{\alpha \ln(\lambda)}}{1 - \sqrt{\alpha \ln(\lambda)}} = \sqrt{\alpha} \frac{|\ln(\lambda)|}{1 + \sqrt{\alpha |\ln(\lambda)|}}. \]  

(51)

But since \( \alpha < 1 \), \( 1 + \sqrt{\alpha |\ln(\lambda)|} < 1 + |\ln(\lambda)| \). Furthermore, for all \( \lambda \leq a < 1 \), by the monotonicity of the function \( t \rightarrow |\ln(t)|/(1 + |\ln(t)|) = -\ln(t)/(1 - \ln(t)) \) on \((0, 1)\), we get that

\[ \frac{|\ln(t)|}{1 + |\ln(t)|} \geq \frac{|\ln(c)|}{1 + |\ln(c)|} \quad \forall t \in (0, a). \]

By applying the above inequality to (51) and taking the square, we get:

\[ \forall \lambda \in (0, c), \quad (1 - \lambda \sqrt{\alpha})^2 \geq M\alpha \quad \text{with} \quad M = \left( \frac{|\ln(a)|}{1 + |\ln(a)|} \right)^2. \]

Whence the following inequality:

\[ \frac{1}{\lambda + (1 - \lambda \sqrt{\alpha})^2} \leq \frac{1}{\lambda + M\alpha}, \]  

(52)

which implies that

\[ \sqrt{\lambda}g_{\alpha}(\lambda) \leq \frac{\lambda^{1/2}}{\lambda + M\alpha}. \]  

(53)

It is rather straightforward to prove that the supremum over \( \lambda \in (0, 1) \) of the right hand side of (53) is of order \( \alpha^{-1/2} \) from which we deduce that

\[ \sup_{\lambda \in (0, a]} \sqrt{\lambda}g_{\alpha}(\lambda) = O\left( \frac{1}{\sqrt{\alpha}} \right). \]  

(54)

\[ \square \]

Proof of Lemma 1. Let \( \lambda \in (0, 1) \). On the one hand, by applying the estimate \((1 - \exp(t)) \geq -t/(1 - t)\) which holds for all \( t < 0 \) to \( t = \sqrt{\alpha \ln(\lambda)} \) and by taking squares, we have:

\[ (1 - \lambda \sqrt{\alpha})^2 \geq \frac{\alpha |\ln(\lambda)|^2}{(1 + \sqrt{\alpha |\ln(\lambda)|})^2}. \]  

(55)

On the other hand, using the estimate \( t^2 \geq (1 - \exp(t))^2 \) valid for all \( t < 0 \) to \( t = \sqrt{\alpha \ln(\lambda)} \), we get

\[ (1 - \lambda \sqrt{\alpha})^2 \leq \alpha |\ln(\lambda)|^2 \]  

(56)

Now, for \( \alpha \leq \lambda < 1 \), \( |\ln(\alpha)| \geq |\ln(\lambda)| \) which implies that \( \sqrt{\alpha} |\ln(\lambda)| \leq \sqrt{\alpha} |\ln(\alpha)| \). Using the estimate \( t^\mu \ln(1/t) \leq \mu \) which is true for all \( t \) in \((0, 1)\) and every positive \( \mu \) to \( t = \lambda \) and \( \mu = 1/2 \), we deduce that

\[ 1 + \sqrt{\alpha} |\ln(\lambda)| \leq 3/2. \]  

(57)
Let $\lambda < c$ then we get

$$h(\lambda) = \frac{1}{\lambda + (4/9)\alpha|\ln(\lambda)|^2}$$

with $\lambda \geq (4/9)\lambda$ to deduce the existence of a root of the function $h$ on $(0, \lambda]$. If $p < 2$, then we get $h(1) = -(2 - p) < 0$, hence the function $h$ vanishes on $(0, 1)$ and there exists a maximizer of the function $h(p, \alpha)$ on $(0, 1)$ which cancels the function $h$. Next, if $p \geq 2$, Let $\lambda < c \leq \exp(2 - p) < 1$, then

$$0 < |\ln(c)| + 2 - p < |\ln(\lambda)| + 2 - p \Rightarrow h(\lambda) < \alpha p|\ln(\lambda)|^2 - \tau_1 \lambda$$

with $\tau_1 = |\ln(c)| + 2 - p > 0$. Since $\lim_{\lambda \to 0} \alpha p|\ln(\lambda)|^2 - \tau_1 \lambda = -\tau_1 \lambda < 0$, we deduce that for $\alpha$ small $h(\lambda) < \alpha p|\ln(\lambda)|^2 - \tau_1 \lambda < 0$. In the case $p \geq 2$, provided that $\alpha$ is small, this proves the existence of a maximizer of the function $\Psi(p, \alpha)$ which cancels the function $h$. Now let us show that for every $\lambda(p, \alpha)$ which vanishes $h$, (33) holds.

$$h(\lambda) = 0 \quad \Rightarrow \quad \alpha = \lambda|\ln(\lambda)|^{-1} \left(\frac{2 - p + |\ln(\lambda)|}{p|\ln(\lambda)|}\right)$$

by monotonicity of the function $t \to (2 - p + t)/(pt)$ (irrespective of the sign of $2 - p$) and $t \to |\ln(\lambda)|$, we get that the function $l(\lambda) = \frac{2 - p + |\ln(\lambda)|}{p|\ln(\lambda)|}$ is monotonic. If $p < 2$, the function $l$ is increasing and we then get that, for all $\lambda \in (0, c]$ with $c < 1$,

$$\frac{1}{p} \leq l(\lambda) \leq l(c).$$

On the other hand, if $p \geq 2$, the function $l$ is decreasing and for $\lambda \in (0, c]$ with $c < \exp(2 - p)$, we get

$$l(c) \leq l(\lambda) \leq 1/p.$$ 

From (60), (61) and (62), we deduce that

$$h(\lambda) = 0 \quad \Rightarrow \quad \alpha \sim \lambda|\ln(\lambda)|^{-1}.$$ 

From [31, Lemma 3.3], we get that

$$\alpha \sim \lambda|\ln(\lambda)|^{-1} \Rightarrow \lambda \sim \alpha|\ln(\alpha)|(1 + o(1)) \quad \text{for} \quad \alpha \to 0.$$ 

This shows that the maximizers $\lambda(p, \alpha)$ of the function $\Psi(p, \alpha)$ satisfies (33). Now let us deduce (34). We have

$$\alpha|\ln(\alpha)|^p \Psi(p, \alpha)(\alpha|\ln(\alpha)|) = \frac{|\ln(\alpha)|^p \times |\ln(\alpha| \ln(\alpha))|^{2-\rho}}{|\ln(\alpha)| + |\ln(\alpha| \ln(\alpha))|^{2}} < |\ln(\alpha)|^p \times |\ln(\alpha| \ln(\alpha))|^{-\rho}$$
With the change of variable \( \varrho = |\ln(\alpha)| \) (i.e. \( \alpha = \exp(-\varrho) \)), we have

\[
|\ln(\alpha)|^{p} \times |\ln(\alpha|ln(\alpha)|)|^{-p} = \frac{\varrho^{p}}{|\ln(\alpha|exp-\varrho|^{p}) = \frac{|-\varrho + \ln(\varrho)|^{p}}{\varrho^{p}} = \frac{(\varrho - \ln(\varrho))^{p}}{\varrho^{p}} \rightarrow 1 \quad \text{as} \quad \varrho \rightarrow \infty.
\]

This proves that

\[
\alpha|\ln(\alpha)|^{p}\Psi_{p,\alpha}(\alpha|ln(\alpha)|) = O(1)
\]

and thus from (33), we deduce that (34) holds.

\[
\square
\]

References

[1] N. Alibaud, P. Maréchal and Y. Saesor, A variational approach to the inversion of truncated Fourier operators, Inverse Problems 25 (2009), no. 4.

[2] M. L. Baart, The use of auto-correlation for pseudorank determination in noisy ill-conditioned linear least-squares problems, IMA J. Numer. Anal. 2 (1982), no. 2, pp 241-247.

[3] F. Bauer and S. Kindermann, Recent results on the quasi-optimality principle, J. Inverse Ill-Posed Probl. 17 (2009), no. 1, pp 518.

[4] F. Bauer and S. Kindermann, The quasi-optimality criterion for classical inverse problems, Inverse Problems 24 (2008), no. 3.

[5] F. Bauer and M. Rei, Regularization independent of the noise level: an analysis of quasi-optimality, Inverse Problems 24 (2008), no. 5.

[6] X. Bonnefond and P. Maréchal, A variational approach to the inversion of some compact operators, Pac. J. Optim. 5 (2009), no. 1, pp 97-110.

[7] H. W. Engl, M. Hanke and A. Neubauer, Regularization of inverse problems, Mathematics and its Applications, 375. Kluwer Academic Publishers Group, Dordrecht, 1996.

[8] G. H. Golub, M. Heath and G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979), no. 2, pp 215-223.

[9] C. W. Groetsch, Inverse problems in the mathematical sciences, Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.

[10] M. Hanke and P. C. Hansen, Regularization methods for large-scale problems, Surveys Math. Indust. 3 (1993), no. 4, pp 253-315.

[11] P. C. Hansen, Regularization Tools version 4.0 for Matlab 7.3, Numer. Algorithms 46 (2007), no. 2, pp 189-194.

[12] B. Hofmann and P. Mathé, Analysis of profile functions for general linear regularization methods, SIAM J. Numer. Anal. 45 (2007), no. 3, pp 1122-1141.

[13] T. Hohage, Regularization of exponentially ill-posed problems, Numer. Funct. Anal. Optim. 21(2000), no. 3-4, pp 439-464.
[14] A. Kirsch, An introduction to the mathematical theory of inverse problems, Applied Mathematical Sciences, 120. Springer-Verlag, New York, 1996.
[15] A. S. Leonov, On the choice of regularization parameters by means of quasi-optimality and ratio criteria, Soviet. Math. Dokl. 19 (1978), no. 3.
[16] O. V. Lepskiǐ, A problem of adaptive estimation in Gaussian white noise, Teor. Veroyatnost. i Primenen. 35 (1990), no. 3, pp 459-470.
[17] A. K. Louis, A unified approach to regularization methods for linear ill-posed problems, Inverse Problems 15 (1999), no. 2, pp 489498.
[18] A. K. Louis and P. Mass, A mollifier method for linear operator equations of the first kind, Inverse Problems 6 (1990), no. 3, pp 427440.
[19] M. A. Lukas, Asymptotic optimality of generalized cross-validation for choosing the regularization parameter, Numer. Math. 66 (1993), no. 1, pp 4166.
[20] B. A. Mair, Tikhonov regularization for finitely and infinitely smoothing operators, SIAM J. Math. Anal. 25 (1994), no. 1, pp 135147.
[21] P. Mathé, Saturation of regularization methods for linear ill-posed problems in Hilbert spaces, SIAM J. Numer. Anal. 42 (2004), no. 3, pp 968973.
[22] P. Mathé and S. V. Pereverzev, Geometry of linear ill-posed problems in variable Hilbert scales, Inverse Problems 19(2003), no. 3, pp 789803.
[23] P. Mathé and B. Hofmann, How general are general source conditions?, Inverse Problems 24 (2008), no. 1.
[24] C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery. Optimal estimation in approximation theory, Proc. Internat. Sympos., Freudenstadt, 1976, Plenum Press (1977), pp 154.
[25] D. A. Murio, The mollification method and the numerical solution of ill-posed problems, A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1993.
[26] M. T. Nair, E. Schock and U. Tautenhahn, Morozov’s discrepancy principle under general source conditions, Z. Anal. Anwendungen 22 (2003), no. 1, pp 199214.
[27] A. Neubauer, On converse and saturation results for regularization methods, Beitrge zur angewandten Analysis und Informatik, Shaker, Aachen, (1994) pp 262270.
[28] A. Neubauer, On converse and saturation results for Tikhonov regularization of linear ill-posed problems, SIAM J. Numer. Anal. 34 (1997), no. 2, pp 517527.
[29] E. Schock, Approximate solution of ill-posed equations: arbitrarily slow convergence vs. superconvergence, Constructive methods for the practical treatment of integral equations (1984), pp 234243.
[30] C. B. Jr. Shaw, Improvement of the resolution of an instrument by numerical solution of an integral equation, J. Math. Anal. Appl. 37 (1972), pp 83112.
[31] U. Tautenhahn, Optimality for ill-posed problems under general source conditions, Numer. Funct. Anal. Optim. 19 (1998), no. 3-4, pp 377398.
[32] A. N. Tikhonov and V. Y. Arsenin, Solutions of ill-posed problems., John Wiley & Sons, 1977.
[33] G. Wahba, Practical approximate solutions to linear operator equations when the data are noisy, SIAM J. Numer. Anal. 14 (1977), no. 4, pp 651667.