Convergence of regularization methods with filter functions for a regularization parameter chosen with GSURE and mildly ill-posed inverse problems

Bruno Sixou

Convergence of regularization methods with filter functions for a regularization parameter chosen with GSURE and mildly ill-posed inverse problems

Bruno Sixou

CREATIS, CNRS UMR5220; Inserm U630; INSA-Lyon; Université Lyon 1; Université de Lyon, F-69621 Villeurbanne Cedex, France

Abstract

In this work, we show that the regularization methods based on filter functions with a regularization parameter chosen with the GSURE principle are convergent for mildly ill-posed inverse problems and under some smoothness source condition. The convergence rate of the methods is not optimal for very ill-posed problems but the efficiency increases with the smoothness of the solution.

Keywords: Inverse problems , SURE , regularization

1. Introduction

In this article, we consider the numerical solution of a linear inverse problem written as:

$$y = Ax \tag{1}$$

where $A : X \to Y$ is a compact injective operator mapping two infinite dimensional separable Hilbert spaces X and Y. We assume that after discretization, the inverse problem is of the form:

$$y^\delta = B_n f + \epsilon \tag{2}$$

where $y_\delta \in \mathbb{R}^n$, $f \in \mathbb{R}^n$ is the true discrete solution, B_n is the discrete approximation of the operator A and ϵ consists of independent and identically distributed (i.i.d) Gaussian errors with variance equal to σ^2, $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$.

Preprint submitted to Journal of Computational and Applied Mathematics April 9, 2020

In order to obtain a stable solution, we consider a regularization functional with a regularization parameter α:

$$J_\alpha(f) = \frac{\|B_n f - y^\delta\|_2^2}{2} + \alpha R(f) \tag{3}$$

where R is the regularizer to add some prior information on the solution. Traditional regularization approaches include Tikhonov or Sobolev regularization [1]. Different regularization terms have been investigated with various convexity and differentiability properties [2].The non-smooth regularizers enforcing sparsity in a suitable domain, e.g. Fourier, wavelet or gradient are very useful in image analysis problems [3, 4] Spectral regularization methods use an appropriate filter on the eigenvalues of the operator that defines the problem . The most commonly used regularization algorithm is the truncated singular value decomposition [1, 5]. The classical Tikhonov regularization can also been understood in the framework of filter regularization. Regularization methods based on filter functions can also be applied in learning, non parametric estimation and problems like those found financial optimization [6, 7, 8].

In the following, we will focuss on Tikhonov regularization with $R(f) = \|f\|_2^2$ but our results can be extended to filter regularization methods. The regularization parameter α has to be chosen carefully[9]. Several rules of choice of the regularization parameters have been investigated in the literature like the L-curve criteria [10] or some discrepancy principles[2, 1]. A Lepskii adaptative procedure has been studied in [11, 12, 13].

Some risk estimators have been proposed to chose the regularization parameters in ill-posed problems. Rules based on the risk estimation with the Stein Unbiased Risk Estimator (SURE) have been investigated [14, 15, 16, 17, 18]. The idea is to select the regularization parameter that minimizes the SURE estimate of the Mean Square Error (MSE). The SURE was originally limited to the Gaussian case and to denoising problems. Some extensions have been studied for multivariate and exponential families [19]. The case of denoising was extented to more general linear operators in Vonesch et al.[20]. For general linear inverse problems, some authors have considered a generalized version GSURE (Generalized Stein Unbiased Risk Estimator) where the risk is measured in the space of the unknown[21, 22, 23, 24].

Yet, there is very few studies about the quality of these risk estimators and it not clear under which conditions the described procedure achieves the best possible reconstruction of the true solution. In papers like [22, 18, 21, 20], the parameter choice rule with GSURE works well and the problems under consideration are not very ill-posed. Some study of the influence of the ill-posedness of the problem and of the degree of smoothness of the unkown solution are presented in Lucka et al.[40] for Tikhonov regularization methods. Some asymptotic convergence results have been obtained as the dimension of the problem tends to infinity. Our motivation is thus to understandd if these estimators are appropriate for very ill-posed problems and to investigate how the quality of the estimators are modified as a function of the ill-posedness of the problem and as a function of the degree of smoothness of the solution.

In this work, we intend to show that a choice of the regularization parameter with the Generalized Sure method ensures the convergence of regularization methods based on filter functions for mildly ill-posed problems. Our method is based on source conditions that measure the smoothness of the solution. Rates of convergence can be calculated with a-priori information on the solution relative to some smoothness class. Moreover, the power decay of the singular values and the ill-posedness of the inverse problem are also taken into account.

The outline of the paper is the following. In the second section, some useful notions are recalled about compact operators, projection methods, spectral projectors. We then present the Generalized SURE estimates. In the next section, we detail the regularization method and the smoothness class assumptions. Then, we estimate the risk as a function of the noise level and of the regularization parameter. We also show that the Tikhonov type regularization method based on filter functions with a regularization parameter chosen with GSURE is a convergent regularization scheme under some restrictive conditions about the ill-posedness of the inverse problem and the smoothness of the solution.

2. Preliminaries

2.1. Compact operators and finite rank approximations

In this work, we will restrict ourselves to linear compact injective operators $A : X \to Y$ between two Hilbert spaces X and Y. For a self-adjoint compact operators C, the spectral theorem [26, 27] shows that there is a complete orthonormal system of eigenvectors $(u_i)_{i\geq 0}$ and eigenvalues $(\rho_i)_{i\geq 0}$ such that, for $f \in X$:

$$Cf = \sum_{i=0}^{\infty} \rho_i < u_i, f > u_i \tag{4}$$

Let Σ the counting measure on $\mathbb{N}$, the former decomposition can be written in the multiplication form $UCf = \rho Uf$, Σ-almost everywhere, with the unitary operator $U : X \rightarrow l_2(\mathbb{N})$ defined by $Uf(i) =< f, u_i >$, $\rho : \mathbb{N} \rightarrow \mathbb{R}$ such that $\rho(i) = \rho_i$ for $i \in \mathbb{N}$. The essential range of ρ is the spectrum $\sigma(C)$ of the operator C.
In the following, the infinite dimensional problem is replaced by a finite dimensional discretized version. The operator A is approximated by a finite dimensional operator $B_n : \mathbb{R}^n \rightarrow \mathbb{R}^n$. We assume that this approximation is obtained by projection methods and that the discrete approximation of the operator can be written as $B_n = Q_n A P_n$ with finite projections and interpolation operators, $P_n : \mathbb{R}^n \rightarrow X$ and $Q_n : Y \rightarrow \mathbb{R}^n$. The discretization errors for the operator A depends on the norms $\epsilon_X = \|A(I - P_n)\|$ and $\epsilon_Y = \|(I - Q_n)A\|$. The discretization errors must be included in the analysis of the SURE and GSURE method. The operator $B_n^t B_n$ is a finite rank self-adjoint operator with closed range. It is possible to consider an orthonormal basis of the range of this operator $(e_i)_{1 \leq i \leq n}$.

$$B_n^t B_n f = \sum_{i=1}^{n} \rho_i(B_n) < e_i, f > e_i \tag{5}$$

where the $\rho_i(B_n)$ are the square of the singular values $\sigma_i(B_n)$ of the operator B_n.

2.2. Singular values and degree of ill-posedness

The decay rate of the positive sequence $\{\sigma_p(A)\}_{p \geq 0}$ of singular values of the operator A towards 0 when $p \rightarrow \infty$ measures the strength of ill-posedness of the inverse problem $Af = y$. This strength is often expressed by a single number η called the degree of ill-posedness. If the decrease of the square of singular values is described by a power law:

$$\rho_p(A) \sim p^{-\eta} \quad p \geq 0 \tag{6}$$

the inverse problem is mildly ill-posed [1]. With increasing η,the numerical difficulties grow. The faster the decay of the singular values, the more severe the ill-posedness of the problem. For exponentially ill-posed problems, there exists an exponent η such that, $\rho_p(A) \sim exp(-p^\eta)$. We recall here some results about the approximate eigenvalues of compact operators [29, 30].

Proposition 2.1. . *Let $A : X \to Y$ be a compact operator and $B_n : X \to Y$ a sequence of operators approximating A. Let λ_0 be an eigenvalue of A of multiplicity m and index ν, and let $(\lambda_n^i)_{i\in I}$ be the eigenvalues of B_n within some small fixed neighborhood of λ_0. Let us assume that:*
1) A and B_n are linear operators for all n
2) $B_n x \to Ax$ as $n \to \infty$, for all $x \in X$
3) The family $(B_n)_{n\geq 0}$ is collectively compact, i.e. $\{B_n x, n \geq 1 \quad \text{and} \quad \|x\| \leq 1\}$ has compact closure in X.

Then, the sum of the multiplicities of the eigenvalues λ_n^i equals the multiplicity of λ_0 and the elements of $(\lambda_n^i)_{i\in I}$ all converge to λ_0 as $n \to \infty$. For some $c > 0$ and sufficiently large n

$$|\lambda_0 - \lambda| \leq c \max\{\|A\phi_j - B_n\phi_j\|^{1/\nu}, 1 \leq j \leq m\} \tag{7}$$

for all $\lambda \in (\lambda_n^i)_{i\in I}$, and for $(\phi_j)_{1\leq j\leq m}$ a basis for $Ker(\lambda_0 - A)^\nu$.

For projection methods as the ones considered for P_n, the hypothesis of Proposition 2.2 are fulfilled [31].

The norm of the approximation $B_n - A$ is bounded by the errors ϵ_X and ϵ_Y:

$$\|A - B_n\| \leq \|A - Q_k A\| + \|Q_k A(I - P_k)\| = \epsilon_X + \epsilon_Y \tag{8}$$

We will thus assume n is large and that the discretization errors ϵ_X and ϵ_Y are very small so that the square of the singular values of B_n satisfies a similar rate of decay than the ones of A:

$$\rho_p(B_n) \sim p^{-\eta} \quad p \geq 0 \tag{9}$$

Under this assumption, we can estimate the function $R(\beta)$ defined as $R(\beta) = \Sigma(\{\rho \geq \beta\})$.

$$R(\beta) = \Sigma(\{\rho \geq \beta\}) = |\{\rho_j \geq \beta\}| \tag{10}$$

$$\sim \sum_{j\leq \beta^{-\eta}} 1 \sim \int_0^{\beta^{-\eta}} dt \sim \beta^{-\eta} \tag{11}$$

2.3. Spectral projectors

We detail in this section some useful results about spectral projectors. In the following, we will use the spectral family of the operator $B_n^t B_n$[32]. We decompose the space $\mathbb{R}^n$ into a direct sum of subspaces V_k in which $B_n^* B_n$ is reduced to the multiplication by the eigenvalue ρ_k of $B_n^* B_n$:

$$\mathbb{R}^n = \oplus V_k \tag{12}$$

We denote P_k the orthogonal projection operator onto V_k and we introduce, for all $\rho \in \mathbb{R}$ the space

$$G_\rho = \oplus_{\rho_k \leq \rho} V_k \tag{13}$$

Let E_ρ the orthogonal projection onto G_ρ The discontinuities of the function $\rho \to E_\rho$ are the eigenvalues ρ_k. In the sense of distribution in $\mathbb{R}$ with values in the space of linear operators defined on $\mathbb{R}^n$, $L(\mathbb{R}^n)$, the derivative of E_ρ can be identified as a measure dE_ρ given by:

$$dE_\rho = \sum_{\rho_k \leq \rho} \delta_{\rho_k} \otimes P_k \tag{14}$$

where $\otimes$ is the tensorial product. Let $\phi_p(t)$ a piecewise continuous function, the operator $\Phi_p(B^* B)$ is then defined by:

$$\Phi_p(B^t B) = \int_0^a \phi_p(\rho) dE_\rho \tag{15}$$

where a is a constant with $\|B^* B\| \leq a$.

Let e_i be an eigenvector for the eigenvalue ρ_i, we have:

$$dE_\rho(e_i) = \delta(\rho - \rho_i) \otimes e_i \tag{16}$$

$$d(E_\rho(e_i), e_i) = d(E_\rho(e_i), E_\rho(e_i)) = \delta(\rho - \rho_i) \tag{17}$$

$$\sum_{i=1}^{n} \int_0^\alpha \rho d(E_\rho(e_i), e_i) == \sum_{i=1}^{n} \int_0^\alpha \rho \delta(\rho - \rho_i) = - \int_0^\alpha \rho dR(\rho) \tag{18}$$

where the second term is the Stieljes integral associated with the decreasing function

$$t \to R(t) = \Sigma(\{\rho \geq t\}) \tag{19}$$

3. The SURE principle and its generalizations

In the following, we denote $f_\alpha(y^\delta)$ the reconstructed solution for the regularization parameter α and the noisy data y^δ. In inverse problems and especially reconstruction problems, the mean squared error (MSE):

$$MSE(\alpha) = \|f - f_\alpha(y^\delta)\|^2/n \tag{20}$$

is a very usual criteria to estimate the quality of the solution of the inverse problem. The following Stein lemma is the basis to obtain estimates of this mean square error [14, 33]. for the standard denoising problem where one observes a realization $y^\delta = (y_i^\delta)_{1\leq i\leq n}$ of an original signal $f = (f_i)_{1\leq i\leq n}$ distorted by an additive white Gaussian noise ϵ of variance σ^2, so that:

$$y = f + \epsilon \tag{21}$$

Lemma 3.1. *Let ϵ an additive white Gaussian noise $\epsilon = (\epsilon_i)_{1\leq i\leq n}$, i.e. $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. Let $F(y) = (f_i(y^\delta))_{1\leq i\leq n}$ be a n-dimensional vector function such that for all i=1...n, $f_i(y^\delta)$ is weakly differentiable with respect to y_i^δ, bounded by some fast increasing function, $|f_i(y)| \leq exp(\|y\|^2/2\beta^2)$ with $\beta > \sigma$, then:*

$$E[\epsilon^t F(y^\delta)] = \sigma^2 E[divF(y^\delta)] \tag{22}$$

where div is the divergence operator in the weak sense.

For Gaussian noise and and denoising problem with $A = I$, the Stein Unbiased Risk Estimate (SURE) is an unbiased estimate of the mean square error[14, 15, 16, 17, 34].

$$E[SURE(\alpha)] = E[MSE(\alpha)] \tag{23}$$

where E is the expectation with respect to the noise ϵ. It is given by:

$$SURE(\alpha) = \|y^\delta - f_\alpha(y^\delta)\|^2/n - \sigma^2 + 2\sigma^2 n^{-1} Tr(J_{f_\alpha(y^\delta)}) \tag{24}$$

where $J_{f_\alpha(y^\delta)}$ is the Jacobian matrix of the reconstructed solution. The evaluation of the $SURE$ requires the knowledge of the noise variance. Similarly, for inverse problems with a direct operator A different from the identity an unbiased estimate of

$$MSE(\alpha) = \|Af - Af_\alpha(y^\delta)\|^2/n \tag{25}$$

is given by:

$$SURE(\alpha) = \|y^\delta - Af_\alpha(y^\delta)\|^2/n - \sigma^2 + 2\sigma^2 n^{-1} Tr(\nabla_{y^\delta} A f_\alpha(y^\delta)) \tag{26}$$

A generalized SURE has been proposed for exponential families[19]. The SURE has been generalized for inverse problems involving a direct operator different from the identity by Vonesch et al [20]. Following [20], we denote B_{inv} the stabilized approximation of the inverse of B_n and $g_\alpha(y) = B_{inv}^t f_\alpha(y)$. The risk given by Eq.20 is rewritten as:

$$MSE(\alpha) = \frac{1}{n}(\|f\|^2 - 2(y^\delta)^t g_\alpha(y^\delta) + 2\epsilon^t g_\alpha(y^\delta) - 2([I - B_{inv}B_n]x)^t f_\alpha(y^\delta) + \|f_\alpha(y^\delta)\|^2) \tag{27}$$

The first term is independent of α. The term $\epsilon^t g_\alpha(y^\alpha)$ can be estimated with the Stein equality under the conditions detailed in the former lemma [14, 33].

$$E[\epsilon^t g_\alpha(y^\delta)] = \sigma^2 E[div g_\alpha(y^\delta)] \tag{28}$$

Neglecting the third term, the following quantity, denoted as $R(\alpha)$, has been proposed to estimate the risk:

$$\begin{aligned} R(\alpha) &= \frac{1}{n}(\|f\|^2 - 2(y^\delta)^t g_\alpha(y^\delta) + 2\sigma^2 div(g_\alpha(y^\delta)) + \|f_\alpha(y^\delta)\|^2) && (29) \\ &= \frac{1}{n}(\|f\|^2 + \|f_\alpha(y^\delta)\|^2 - 2(y^\delta)^t B_{inv}^t f_\alpha(y^\delta) + 2\sigma^2 div(g_\alpha(y^\delta))) && (30) \end{aligned}$$

In [34] a weighted mean square error (WMSE) measure has been proposed to test the accuracy of the reconstruction:

$$WMSE(\alpha) = n^{-1}\|y^\delta - B_n f_\alpha(y^\delta)\|_W^2 \tag{31}$$

where W is a positive definite symmetric weighting matrix. The matrix W is chosen to counterbalance the effects of the direct operator B_n. With the choice $W = B_n^+ = B_n^t(B_n B_n^t)^+$, where M^+ denotes the pseudo-inverse of M, and under the assumptions of the Stein lemma, the random variable Generalized Sure $GSURE(\alpha)$ is an unbiased estimator of $WMSE(\alpha)$[21, 22, 23, 24]:

$$\begin{aligned} GSURE(\alpha) &= n^{-1}\|y^\delta - B_n f_\alpha(y^\delta)\|_W^2 - \sigma^2 n^{-1} tr((B_n B_n^t)^+) \\ &\quad + 2\sigma^2 n^{-1} tr((B_n B_n^t)^+ \nabla_{y^\delta} B_n f_\alpha(y^\delta)) && (32) \\ &= n^{-1}\|f_{ML}(y^\delta) - f_\alpha(y^\delta)\|_W^2 - \sigma^2 n^{-1} tr((B_n B_n^t)^+) \\ &\quad + 2\sigma^2 n^{-1} tr((B_n B_n^t)^+ \nabla_{y^\delta} B_n f_\alpha(y^\delta)) && (33) \end{aligned}$$

where f_{ML} is the maximum likehood estimate and $J_{f_\alpha(y^\delta)}$ is the Jacobian matrix of the reconstructed solution.

With the choice, $W = B_{inv}^t B_{inv}$, $GSURE(\alpha)$ can be rewritten:

$$\begin{aligned} GSURE(\alpha) = n^{-1}\|B_{inv}(y^\delta - B_n f_\alpha(y^\delta))\|^2 - \sigma^2 n^{-1} tr(B_{inv}^t B_{inv}) \\ +2\sigma^2 n^{-1} tr(B_{inv}^t B_{inv} B_n J_{f_\alpha(y^\delta)}) \end{aligned} \tag{34}$$

With $f_\alpha(y^\delta) = B_{inv} y^\delta$, we obtain:

$$\begin{aligned} GSURE(\alpha) = n^{-1}\|B_{inv}(y^\delta - B_n f_\alpha(y^\delta))\|^2 - \sigma^2 n^{-1} tr(B_{inv}^t B_{inv}) \\ +2\sigma^2 n^{-1} tr(B_{inv}^t B_{inv} B_n B_{inv}) \end{aligned} \tag{35}$$

Some asymptotic properties of the risk estimators have been sudied in [40] as $n \to \infty$ based on a singular value decomposition of the direct operator but the smoothness of the solution in relation with the poperties of the operator have not been considered. We show in the following that chosing the regularization parameter α with the minimum of $R(\alpha)$ or $GSURE(\alpha)$ given by Eq.35 leads to a convergent regularization method.

4. Regularization method and smoothness class

It is well-known that the convergence rate for regularized solutions of inverse problems depends on the type of regularization scheme but also on the smoothness class for the true solution. In this section, we detail the regularized estimators used to evaluate the estimators of the risk and the assumption on the smoothness of the ground truth.

4.1. Regularized estimators

In the following, we will use the notations of [12]. Let $\phi_\alpha : \sigma(A^t A) \to \mathbb{R}$ is a filter function defined on the spectra $\sigma(A^t A)$ of the operator $A^t A$ approximating the function $t \to 1/t$, and which is parametrized by a regularization parameter α. Based on this filter function, and on the singular value decompostion of the operator $A^t A$, it is possible to construct regularized estimators with regularization methods of the form:

$$f_\alpha = \Phi_\alpha(A^t A) A^t y^\delta \tag{36}$$

With the discretization of the operator A, the approximate solution is calculated as:

$$f_\alpha = \Phi_\alpha(B_n^t B_n) B_n^t y^\delta. \tag{37}$$

The regularized inverse will be defined as $B_{inv} = \Phi_\alpha(B_n^t B_n) B_n^t$.

Classical assumption about the filter function ϕ_α can be found in [35, 36]. We assume that there exist some positive constants C_1,C_2 such that:

$$\lim_{\alpha \to 0} \phi_\alpha(t) = 1/t \tag{38}$$

$$\sup_{t \in \sigma(A^t A)} |t^{1/2} \phi_\alpha(t)| \le C_1/\sqrt{\alpha} \quad 0 < \alpha < \|A^t A\| \tag{39}$$

$$\sup_{t \in \sigma(A^t A)} |1 - t\phi_\alpha(t)| \le C_2 \quad 0 < \alpha < \|A^t A\|, \tag{40}$$

$$0 \le \phi_\alpha(t) \le 1/\alpha \quad \forall t \in \sigma(A^t A) \tag{41}$$

In the following, we will use the notation $A_1 \asymp A_2$ if there are two positive constants c_1 and c_2 such that:

$$c_1 A_2 \le A_1 \le c_2 A_2 \tag{42}$$

We also assume that:

$$\sup_{t \ge 0} |(1 - t\phi_\alpha(t))\Lambda(t)| \asymp \Lambda(\alpha) \quad 0 < \alpha < \|A^t A\| \tag{43}$$

for the index functions associated to the smoothness class, i.e $\Lambda(\alpha)$ is not only an upper bound of $\sup_{t \ge 0} |(1 - t\phi_\alpha(t))\Lambda(t)|$ but also a sharp estimate of this term. Several regularization methods can be described with these type of filter functions. For Tikhonov regularization $\phi_\alpha(t) = (\alpha + t)^{-1}$, for iterated Tikhonov regularization, $\phi_\alpha(t) = (1 - (\alpha/(t+\alpha))^m)/t$, for spectral cut-off $\phi_\alpha(t) = 1/t$ for $t \ge \alpha$, and $\phi_\alpha(t) = 0$ for $t \le \alpha$.

4.2. Smoothness classes

The convergence rate for the reconstruction methods is determined by some a priori assumption on the exact solution. Following [26], we will measure the smoothness of the function f relative to the smoothing properties of A with a source condtion. We assume there exists $w \in X$, and $T > 0$ such that:

$$f = \Lambda(A^t A) w \quad \|w\| \le T \tag{44}$$

with $\Lambda : \sigma(A^tA) \to [0, \infty[$ a continuous, strictly increasing index function with $\Lambda(0) = 0$. The error estimates are different under different assumptions on the index function. For exponentially ill-posed problems and infinitely smoothing operators, a logarithmic souce conditions is used, $\Lambda(t) = (-\ln t)^{-\nu}$ with $\nu > 0$ [37]. A common assumption is $\Lambda(t) = t^\nu$ with $\nu > 0$ for finitely smoothing operators. In this work, we consider this type of mildly ill-posed problem. For this smoothness class and exact data y, using Eq.40, we see that the error for non noisy data is bounded by [12] :

$$\|\Phi_\alpha(A^tA)A^ty - f\| \leq \rho \sup_{t\in\sigma(A^*A)} |(\phi_\alpha((t)t-1)\Lambda(t)| \leq C\Lambda(\alpha) \tag{45}$$

From that we can infer, that there is a positive constant C_1 such that:

$$|\|\Lambda(A^tA)w\|^2 - \|\Phi_\alpha(A^tA)A^tA\Lambda(A^tA)w\|^2| \leq C_1\Lambda(\alpha) \tag{46}$$

5. Estimate of the risk *R* and of GSURE

In this section, we will estimate successively the risk R and the $GSURE$.

5.1. Estimate of the risk R

In order to rewritte the risk R, we first separate the estimate of f_α for the non noisy data and the term originating from the noise. We set $f_\alpha^1 = \Phi_\alpha(B_n^tB_n)B_n^tB_nf$ and $f_\alpha^2 = \Phi_\alpha(B_n^tB_n)B_n^tB_n\epsilon$, with $f_\alpha = f_\alpha^1 + f_\alpha^2$. With $B_{inv} = \Phi_\alpha(B_n^tB_n)B_n^t$, the second term of $R(\alpha)$ can be written as:

$$\begin{aligned}(y^\delta)^tB_{inv}^tf_\alpha(y^\delta) &= (B_nf+\epsilon)^tB_n\Phi_\alpha(B_n^tB_n)[\Phi_\alpha(B_n^tB_n)B_n^t(B_nf+\epsilon)] \\ &= < f_\alpha^1, f_\alpha^1 > + < f_\alpha^2, f_\alpha^2 > +2 < f_\alpha^1, f_\alpha^2 > \end{aligned} \tag{47}$$

$R(\alpha)$ can thus be rewritten as:

$$R(\alpha) = \frac{1}{n}(\|f\|^2 - \|f_\alpha(y^\delta)\|^2 + 2\sigma^2 div(g_\alpha(y^\delta))) \tag{48}$$
$$= \frac{1}{n}(\|f\|^2 - \|f_\alpha^1(y^\delta)\|^2 - \|f_\alpha^2(y^\delta)\|^2 - 2 < f_\alpha^1(y^\delta), f_\alpha^2(y^\delta) > +2\sigma^2 div(g_\alpha(y^\delta))$$

Estimation of $\|f\|^2 - \|f_\alpha^1(y^\delta)\|^2$

Following Mathe et al. [12, 38, 39] , for f in the smoothness class defined by

Λ, we can bound the noise free term:

$$\begin{aligned}\|f-\Phi_\alpha(B_n^tB_n)B_n^tQAf\| &= \|f-\Phi_\alpha(B_n^tB_n)B_n^tB_nf\|+\|\Phi_\alpha(B_n^tB_n)B_n^t(B_n-Q_nA)f\| \\ &\le C_1\Lambda(\alpha)+C_2\|(I-\Phi_\alpha(B_n^tB_n)B_n^tB_n)(\Lambda(A^*A)-\Lambda(B_n^*B_n)\| \\ &\quad +\frac{C_3}{\sqrt{\alpha}}\|(B_n-Q_nA)f\| \quad (49) \\ &\le C_1\Lambda(\alpha)+C_2\|(\Lambda(A^tA)-\Lambda(B_n^tB_n)\| \\ &\quad +\frac{C_3}{\sqrt{\alpha}}\|(B_n-Q_nA)f\| \quad (50)\end{aligned}$$

where we have used Eq.38 and Eq.41.
The bound on the error free term depends on the class of the index function Λ. We restrict in this work to functions $\Lambda : t \to t^\mu$ with $0 < \mu \le 1$. This type of index function is operator monotone and satisfies the following bound [12, 38, 39]:

$$\|\Lambda(A^tA)-\Lambda(B^tB)\| \le \Lambda(\|A^tA-B^tB\|) \quad (51)$$

Different estimates of the error can be obtained for different classes of index functions. If the index function Λ belongs to the class $\mathcal{F}_0$ with:

$$\mathcal{F}_0 = \{\phi, \phi^2 \text{ is concave}\} \quad (52)$$

then the following bound on the error holds true[12, 38, 39]:

$$\|f-\phi_\alpha(B_n^tB_n)B_n^tQ_nAf\| \le C_1(\Lambda(\alpha)+\Lambda(\epsilon_X^2)+C_2\Lambda(\epsilon_Y^2)+C_3\frac{\Lambda(\epsilon_Y^2)}{\sqrt{\alpha}} \quad (53)$$

Similar bounds can be found for different index functions. For the class $\mathcal{F}_{1/2}$ defined by:

$$\mathcal{F}_{1/2} = \{\phi, \phi \le c\sqrt{t} \quad \text{for some} \quad c>0\} \quad (54)$$

it can be shown that:

$$\|f-\phi_\alpha(B_n^*B_n)B_n^*Q_nAf\| \le C_1(\Lambda(\alpha)+\Lambda(\epsilon_X^2))+C_2\Lambda(\epsilon_Y^2)+C_3\epsilon_Y+C_4\frac{\epsilon_Y^2}{\sqrt{\alpha}} \quad (55)$$

For the class

$$\mathcal{F}_1 = \{\phi, \phi=\psi\theta, \psi \in \mathcal{F}_0, \theta \quad \textit{Lipschitz}\} \quad (56)$$

the error bound is given by:

$$\|f-\phi_\alpha(B^*B)B^*QAf\| \le C_1(\Lambda(\alpha)+\theta(\alpha)\psi(\epsilon_X^2)+\theta(\alpha)\epsilon_Y)+C_2(\epsilon_X^2+2\alpha\epsilon_Y)+C_3\epsilon_Y+C_4\frac{\epsilon_Y^2}{\sqrt{\alpha}}) \quad (57)$$

As explained in[12, 38, 39], it is possible to chose a high discretization level ϵ_X and ϵ_Y such that the dominant term in the error is $\Lambda(\alpha)$. We will make this assumption in the following. To conclude, with Eq.43 we obtain that there is a positive constant C such that:

$$\|f\|^2 - \|f_\alpha^1(y^\delta)\|^2 \asymp C\Lambda(\alpha) \tag{58}$$

Estimation of $2\sigma^2 div(g_\alpha(y^\delta))$
With

$$f_\alpha(y^\delta) = \Phi_\alpha(B_n^t B_n)B_n^t(y^\delta) = B_{inv}y^\delta \tag{59}$$

and $g_\alpha(y^\delta) = B_{inv}^t f_\alpha(y^\delta)$, $B_{inv}^t = B_n\Phi_\alpha(B_n^t B_n)^t = B_n\Phi_\alpha(B_n^t B_n)$ we obtain

$$div(g_\alpha(y)) = Tr[B_n\Phi_\alpha(B_n^t B_n)^t\Phi_\alpha(B_n^t B_n)B_n^t] \tag{60}$$
$$= \textstyle\sum_{1\leq i\leq n} \|\Phi_\alpha(B_n^t B_n)B_n^t e_i\|^2 \tag{61}$$

Using the spectral projectors, and for ϕ_α the filter function for the Tikhonov regularization, we get:

$$\begin{aligned}\|\Phi_\alpha(B_n^* B_n)B_n^* e_i\|^2 &= \int_0^\infty \phi_\alpha^2(\rho)\rho d\|E_\rho(e_i)\|^2 \\ &= \int_0^\infty \frac{\rho}{(\rho+\alpha)^2} d\|E_\rho(e_i)\|^2 \end{aligned}\tag{62}$$

We will split this integral in Eq.62 into two terms.
For $0 < \rho < \alpha$, we have $1/4 \leq \frac{1}{(1+\rho/\alpha)^2} \leq 1$ and thus:

$$I_1 = \sum_i \int_{\rho<\alpha} \frac{\rho}{(\rho+\alpha)^2} d\|E_\rho(e_i)\|^2 = \frac{1}{\alpha^2}\sum_i \int_{\rho<\alpha} \frac{\rho}{(1+\rho/\alpha)^2} d\|E_\rho(e_i)\|^2$$
$$\asymp \frac{1}{\alpha^2}\sum_i \int_{\rho<\alpha} \rho d\|E_\rho(e_i)\|^2 \tag{63}$$
$$\asymp \frac{1}{\alpha^2}\sum_i \int_{\rho<\alpha} \rho\delta(\rho-\rho_i) \tag{64}$$

We can reformulate it with the Stieljes integral:

$$I_1 \asymp -\frac{1}{\alpha^2}\int_0^{\rho<\alpha} \beta dR(\beta) \tag{65}$$

Let us assume that R is smooth with $R(\beta) \sim \beta^{-\eta}$, we obtain; $I_1 \asymp \frac{1}{\alpha^2}(\alpha)^{1-\eta} \asymp \alpha^{-1-\eta}$
The same result can be obtained assuming Eq.41 is satisfied and for more general filter functions.

For $\rho > \alpha$, $1/4 \leq \frac{1}{(1+\alpha/\rho)^2} \leq 1$ and thus the second term can be written:

$$I_2 = \sum_i \int_{\rho>\alpha} \frac{\rho}{(\rho+\alpha)^2} d\|E_\rho(e_i)\|^2 \tag{66}$$

$$\sim \sum_i \int_{\rho>\alpha} \frac{1}{\rho(1+(\alpha/\rho))^2} d\|E_\rho(e_i)\|^2 \tag{67}$$

$$\tag{68}$$

Similarly, the term I_2 can be written as:

$$I_2 \asymp -\int_{\rho>\alpha} \frac{1}{\beta} dR(\beta) \asymp \alpha^{-1-\eta} \tag{69}$$

To conclude, we obtain:

$$2\sigma^2 div(g_\alpha(y^\delta)) \asymp \sigma^2 \alpha^{-1-\eta} \tag{70}$$

Estimation of $< f_\alpha^1, f_\alpha^2 >$

We can also estimate the scalar product $< f_\alpha^1, f_\alpha^2 >$

$$< f_\alpha^1, f_\alpha^2 >=< \Phi_\alpha(B_n^t B_n) B_n^t B_n f, \Phi_\alpha(B_n^t B_n) B_n^t \epsilon > \tag{71}$$

$$=< \Phi_\alpha(B_n^t B_n) B_n^t B_n \Lambda(A^t A) w, \Phi_\alpha(B_n^t B_n) B_n^t \epsilon > \tag{72}$$

$$= \int \phi_\alpha(\rho)^2 \Lambda(\rho) \rho^{3/2} d < E_\rho w, \epsilon > \tag{73}$$

The random measure $d < E_\rho w, \epsilon >$ can be written

$$dE_\rho w = \sum_{i=1}^n \delta(\rho - \rho_i) \otimes < e_i, w > e_i \tag{74}$$

$$d < E_\rho w, \epsilon >= \sum_{i=1}^n \delta(\rho - \rho_i) \otimes < e_i, w >< e_i, \epsilon > \tag{75}$$

$< f_\alpha^1, f_\alpha^2 >$ is the linear combination of gaussian variables and it is a Gaussian variable with $E[< f_\alpha^1, f_\alpha^2 >] = 0$ and:

$$Var(< f_\alpha^1, f_\alpha^2 >) = \sigma^2 \int \phi_\alpha(\rho)^2 \Lambda(\rho) \rho^{3/2} \sum_{i=1}^{n} \delta(\rho - \rho_i) < e_i, w > \asymp \sigma^2$$

With a simple concentration inequality like the Chebyshev's inequality, the probability is very small that $< f_\alpha^1, f_\alpha^2 >$ is much larger that is expected value $E[< f_\alpha^1, f_\alpha^2 >] = 0$.

Estimation of $\|f_\alpha^2\|^2$

Similarly, we have:

$$\|f_\alpha^2\|^2 = \int \phi_\alpha(\rho)^2 \rho^2 d < E_\rho \epsilon, E_\rho \epsilon > \tag{76}$$

The component $(< e_i, \epsilon >)_{1 \leq i \leq n}$ are i.i.d Gaussian with the law $\mathcal{N}(0, \sigma^2)$.

$$E_\rho \epsilon = \sum_{\lambda_i \leq \rho} < e_i, \epsilon > e_i \tag{77}$$

and

$$\|E_\rho \epsilon\|^2 = \sum_{\lambda_n \leq \rho} | < e_n, \epsilon > |^2 \tag{78}$$

This function is an increasing right continuous function and defines a random Stieljes measure.

$$E_{\rho_i} \epsilon - E_{\rho_{i-1}} \epsilon = < e_i, \epsilon > e_i \tag{79}$$

and thus, we obtain the random measure:

$$dE_\rho \epsilon = \sum_{i=1}^{n} \delta(\rho - \rho_i) \otimes < e_i, \epsilon > e_i \tag{80}$$

where $\otimes$ is the tensorial product of distributions

$$d < E_\rho \epsilon, E_\rho \epsilon > = \sum_{i=1}^{n} \delta(\rho - \rho_i) | < e_i, \epsilon > |^2 \tag{81}$$

The average measure dm is given by:

$$dm(\rho) = \sum_{i=1}^{n} \delta(\rho - \rho_i)\sigma^2 = -\sigma^2 dR(\rho) \quad (82)$$

The integral defining $E[\|f_\alpha^2\|^2]$ can be splitted in two terms I_1 and I_2:

$$I_1 = E[\int_0^\alpha \phi_\alpha(\rho)^2\rho^2 d < E_\rho\epsilon, E_\rho\epsilon >] \quad (83)$$

$$= \sigma^2 \int_0^\alpha \phi_\alpha(\rho)^2\rho^2 \sum_{i=1}^{n} \delta(\rho - \rho_i) \asymp -\frac{\sigma^2}{\alpha^2}\int_0^\alpha \rho^2 dR(\rho) \asymp \sigma^2\alpha^{-\eta} \quad (84)$$

$$I_2 = E[\int_\alpha^\infty \phi_\alpha(\rho)^2\rho^2 d < E_\rho\epsilon, E_\rho\epsilon >] \asymp \sigma^2 \int_\alpha^\infty \sum_{i=1}^{n} \delta(\rho - \rho_i) \quad (85)$$

$$\asymp \sigma^2 \int_\alpha^\infty d\Sigma(\rho) \quad (86)$$

$$\asymp -\sigma^2 \int_\alpha^\infty dR(\rho) \asymp \sigma^2\alpha^{-\eta} \quad (87)$$

To conclude, we obtain:

$$E[\|f_\alpha^2\|^2] \asymp \sigma^2\alpha^{-\eta} \quad (88)$$

Using Eq.58 , Eq.70 , Eq.88:

$$E[R(\alpha)] \sim \frac{1}{n}(\alpha^\mu + C_1\sigma^2\alpha^{-1-\eta} + C_2\sigma^2\alpha^{-\eta}) \quad (89)$$

The third term being negligible as $\alpha \to 0$, the mininimization of $R(\alpha)$ gives $\alpha \sim \sigma^{2/(1+\eta+\mu})$ and the average risk scales as $R(\alpha) \sim \sigma^{2\mu/(1+\eta+\mu})$.

5.2. Estimation of GSURE

We start from Eq.35 to estimate $GSURE$:

$$Tr(B_{inv}B_{inv}^t) = Tr[B_n\Phi_\alpha B_n^t B_n)^t\Phi_\alpha(B_n^t B_n)B_n^t] \quad (90)$$

$$= \sum_{1\leq i\leq n} \|\Phi_\alpha(B_n^t B_n)B_n^t e_i\|^2 \quad (91)$$

$$\asymp \alpha^{-1-\eta} \quad (92)$$

with Eq.70.
With the spectral projectors, the power law of operators are well-defined and we have:

$$I = Tr(B_{inv}B_{inv}^t B_n B_{inv}) = Tr[(B_n B_n^t)^2 \Phi_\alpha (B_n^t B_n)^3[\tag{93}$$

$$= \sum_{1\leq i\leq n} \|\Phi_\alpha^{3/2}(B_n^t B_n) B_n B_n^t e_i\|^2 \tag{94}$$

$$= \sum_i \int_0^\infty \rho^2 \phi_\alpha^3(\rho) d\|E_\rho(e_i)\|^2 \tag{95}$$

We split the integral into two parts:

$$1_1 = \sum_i \int_0^\alpha \rho^2 \phi_\alpha^3(\rho) d\|E_\rho(e_i)\|^2 \asymp -\frac{1}{\alpha^3}\int_0^\alpha \rho^2 dR(\rho) \asymp -\frac{1}{\alpha^3}\alpha^{2-\eta} \asymp \alpha^{-1-\eta}$$

$$1_2 = \sum_i \int_\alpha^\infty \rho^2 \phi_\alpha^3(\rho) d\|E_\rho(e_i)\|^2 \asymp \sum_i \int_\alpha^\infty \frac{1}{\rho} d\|E_\rho(e_i)\|^2 \tag{96}$$

$$\asymp -\int_\alpha^\infty \frac{1}{\rho} dR(\rho) \asymp \alpha^{-1-\eta} \tag{97}$$

It follows:

$$Tr(B_{inv}B_{inv}^t B_n B_{inv}) \asymp \alpha^{-1-\eta} \tag{98}$$

Using the decomposition $f_\alpha = f_\alpha^1 + f_\alpha^2$, we obtain as $\alpha \to 0$ and for a high discretization level:

$$\|f_\alpha - B_{inv}B_n f_\alpha\| \asymp \|f_\alpha^1 - B_{inv}B_n f_\alpha^1\| = \|(I - B_{inv}B_n) f_\alpha^1\| \tag{99}$$

$$= \|(I - \Phi_\alpha(B_n^t B_n) B_n^t B_n)\Phi_\alpha(B_n^t B_n) B_n^t B_n f\| \asymp \Lambda(\alpha) \tag{100}$$

with Eq.43.
Using Eq.92 , Eq.98 , Eq.100, GSURE can be estimated as:

$$GSURE(\alpha) \sim \frac{1}{n}(\alpha^{2\mu} + C_1 \sigma^2 \alpha^{-1-\eta}) \tag{101}$$

The mininimization of $GSURE$ gives $\alpha \sim \sigma^{2/(1+\eta+2\mu)}$ and $GSURE \sim \sigma^{4\mu/(1+\eta+2\mu)}$.

6. Discussion

The convergence of the Tikhonov regularization method is obtained for $\alpha \to 0$, and $\frac{\sigma^2}{\alpha} \to 0$[2]. The choice $\alpha \sim \sigma^{2/(1+\eta+2\mu)}$ obtained with the minimization of GSURE or R ensures thus the convergence of the Tikhonov regularization for a regularization parameter chosen with these estimators.

It is interesting to compare the convergence rate obtained with the minimization of $R(\alpha)$ or $GSURE(\alpha)$ with the optimal convergence rates for projection methods. Let S be a given reconstruction method, the worst case error at f for a Gaussian noise ϵ with variance σ^2 is then given by:

$$e(f, S, \sigma^2) = \sup_{\epsilon} \|f - S(y_\delta)\| \tag{102}$$

The best possible accuracy is defined by the minimization over all the numerical numerical methods:

$$e(f, \sigma^2) = \inf_{S:Y \to X} e(f, S, \sigma^2) \tag{103}$$

We are interested by the asymptotic behavior of $e(f, \sigma^2)$ as $\sigma \to 0$.

The optimal order of approximation, for exact solutions satisfying the source condition of Eq.44 has been studied in Mathe et al.[39] for projections methods.

Proposition 6.1. *Let Λ an index function such that $\Theta(t) = \sqrt{t}\Lambda(t)$ is strictly increasing, with $\Theta(t) \to 0$ as $t \to 0$, such that there is a positive constant c for which $\Lambda(2t) \leq c\Lambda(t)$, for $0 < t < a$, and such that $t \to \Lambda^2((\Theta^2)^{-1})$ is concave, then the best possible accuracy is of the order of $\Lambda(\Theta^{-1}(\sigma^2))$.*

The former conditions are satisfied for $\Lambda(t) = t^\mu$ considered in this work. For finite projections methods, the best possible accuracy is thus of the order of $\Lambda(\Theta^{-1}(\sigma^2))$ where $\Theta(t) = \sqrt{t}\Lambda(t)$. For index functions of the type $\Lambda(t) = t^\mu$, we obtain that the best possible accuracy is of order $\sigma^{2\mu/(1+2\mu)}$. Since $2\mu/(1+\eta+2\mu) < 2\mu/(1+2\mu)$, we can conclude that a regularization parameter chosen with the GSURE or R is not optimal. As η increases, the ill-posedness increases and the efficiency of the method of choice of the parameter decreases. If the condition number of B_n increases, smaller singular values are added and it is expected that the GSURE is less efficient.

7. Numerical Experiments

In order to illustrate the former results, we consider the following linear inverse problem which has been studied in detail in [40]. The linear convolution operator A is mapping a function $f : [-1/2, 1/2] \to \mathbb{R}$ to a function $y : [-1/2, 1/2[\to \mathbb{R}$ with a compacted supported kernel with a parameter $0 < l < 1/2$ defined as:

$$k_l = \frac{1}{N_l} \begin{cases} exp(-\frac{1}{1-t^2/l^2}) & \text{if } \ |t| < l \\ 0 & \text{if } \ l \le |t| \le 1/2 \end{cases} \tag{104}$$

with $N_l = \int_{-l}^{l} exp(-\frac{1}{1-t^2/l^2})dt$. The function f used is the peak $f_1 = \delta(t - b_1)$ with $b_1 = \frac{1}{\sqrt{26}} - 0.5$ or the smooth function $f_2(t) = \exp(-\frac{t^2}{4})$.

For a discretization level n for the domain and the range of the operator A, we denote $E_i^n = [\frac{i-1}{n} - \frac{1}{2}, \frac{i}{n} - \frac{1}{2}]$ for $1 \le i \le n$, and $\psi_i^n(t) = \sqrt{n} 1_{E_i^n}(t)$ the associated orthonormal basis. Some Gaussian noise with standard deviation σ was added to the data y. From the singular value decomposition of the convolution operator, the GSURE can be easily estimated [40]. Let (γ_i) be the singular values of the operator A, for the noise level σ and a regularization parameter α, we can compute:

$$GSURE(\alpha) = \sum_{i=1}^{n} (\frac{1}{\gamma_i} - \frac{\gamma_i}{\gamma_i^2 + \alpha})^2 (y^\delta)^2 - \sigma^2 \sum_{i=1}^{n} \frac{1}{\gamma_i^2} + 2\sigma^2 \sum_{i=1}^{n} \frac{1}{\gamma_i^2 + \alpha} \tag{105}$$

The optimal parameter $\alpha_{opt,GSURE}$ is obtained as:

$$\alpha_{opt,GSURE} = \operatorname*{argmin}_{\alpha \ge 0} GSURE(\alpha) \tag{106}$$

In [22], the authors have shown that GSURE is differentiable in a weak sense, and have performed quasi-Newton optimization based on the first-order information given by an unbiased estimator of the gradient of the risk. Here, we have evaluated the risk for different values of the regularization parameter α and plotted its evolution to determine the optimal value of α. The minimum of GSURE can be found efficiently and the computational cost is not two high for a scalar real-valued parameter.

The discretized operator and functions are written as follows:

$$A_{i,j} =< \psi_i^n, A\psi_j^n > n \int_{E_i^n} \int_{E_j^n} k_l(s-t) ds dt \tag{107}$$

$$< \psi_i^n, x >= \sqrt{n} \int_{E_i^n} f(t)dt \tag{108}$$

The parameters n and l control the ill-posedness of the inverse problem. The ill-posedness and the condition number increase with higher n values and lower l values. For examples, the parameters $n = 128$, $l = 0.1$, with a condition number 4.3 e+4, corresponds to a very ill-posed problem and the choice $n = 32$, $l = 0.02$ to a less ill-posed problem,with condition number 1.75.

In order to illustrate the efficiency of the GSURE we have evaluated the optimal parameter for the GSURE and for the l_2 error $\|f - f_\alpha(y^\delta)\|_2$ for the former inverse problem with different ill-posedness degrees related to the discretization level n and to the value of l, with different smoothness of the solution and with different noise levels σ. The $GSURE$ and the error have been evaluated on a fine logarithmical grid , where $log(\alpha)$ is increased linearly with a step size 0.01 between a lower bound α_{min} and $\alpha_{min} + 10$. Typical curves for the error and GSURE are displayed in Figure 1 and 2.

$Ne = 10000$ samples of the noise have been drawn. Typical joint distributions for the optimum values of α for the reconstruction error and the GSURE are displayed on Figure 3, 4 and 5. The Figure 3 corresponds to values of n and l for a moderately ill-posed problem and the peak f_1 and Figure 4 to values of these parameters for a very ill-posed problem and the same function. Figure 5 displays the results obtained for a very ill-posed problem and a smooth function f_2.

As displayed on Figure 3, for n and l values corresponding to a low condition number, the regularization parameters obtained with the minimization of the error or with GSURE are in good agreement and a peak is obtained on the diagonal of the plot. As shown on Figure 4, for a very ill-posed problem and a non smooth solution, the optimal value of α derived from the minimization of GSURE leads to large l_2 errors and the optimal values of the regularization parameter for the minimal l_2 is much larger. It can be seen on Figure 5 that for a smooth function f_2 and a very ill-posed problem the deviation between the two parameters obtained with GSURE of the l_2 errors is also large. These results illustrate the former results and suggest that GSURE is a good estimate of the optimal regularization parameter. Yet, it is not optimal for very ill-conditioned problems and it underestimates the regularization parameters in this case.

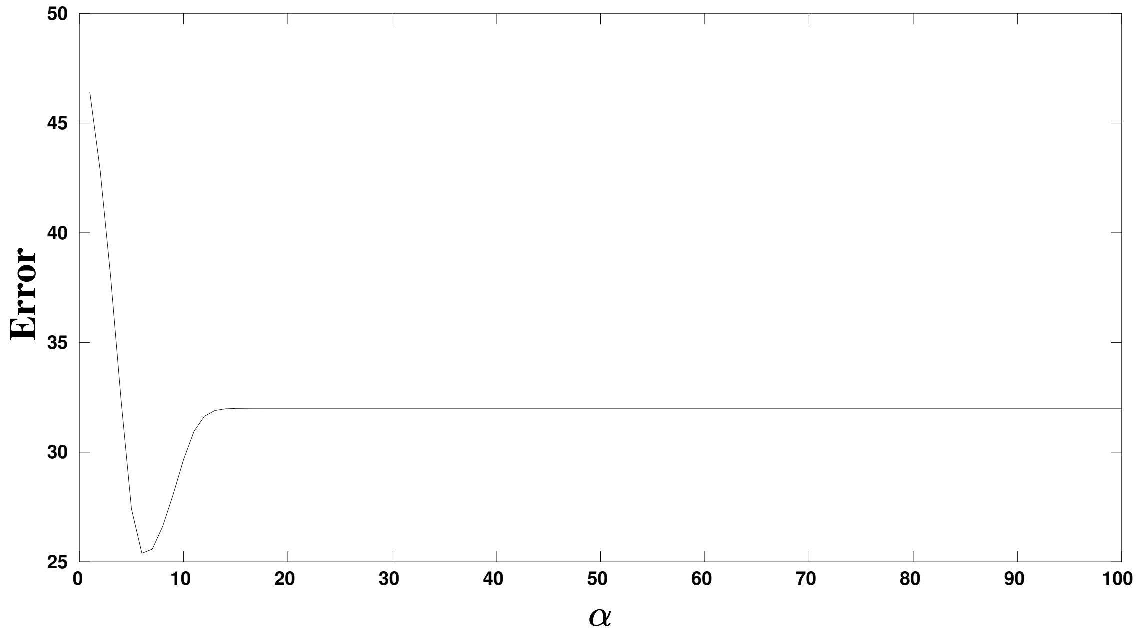

Figure 1: Typical evolution of the l_2 error term for the parameters n=32, l=0.01 and $\sigma = 0.1$.

8. Conclusion

We have estimated the risk R or $GSURE$ for regularization methods based on filter functions such as the Tikhonov regularization. The results are obtained with spectral projectors and are based on some restrictive conditions about the ill-posedness of the inverse problem and the smoothness of the solution.

We have shown that the choice of the regularization parameter for regularization with filter functions based on the minimization of the risk R or of the $GSURE$ estimate gives a convergent regularization method for mildly ill-posed problems. This result is obtained under the assumption that the discretization errors can be neglected. Yet, the best possible accuracy is not achieved and the regularization method is not optimal for the smoothness class defined by the exponent μ. With growing μ, the spaces of functions satisfying the source condition contains smoother elements and the quality of the regularization with the parameter chosen with the GSURE principle increases but the method is still not optimal. Moreover, GSURE is not optimal for very ill-conditioned problems.

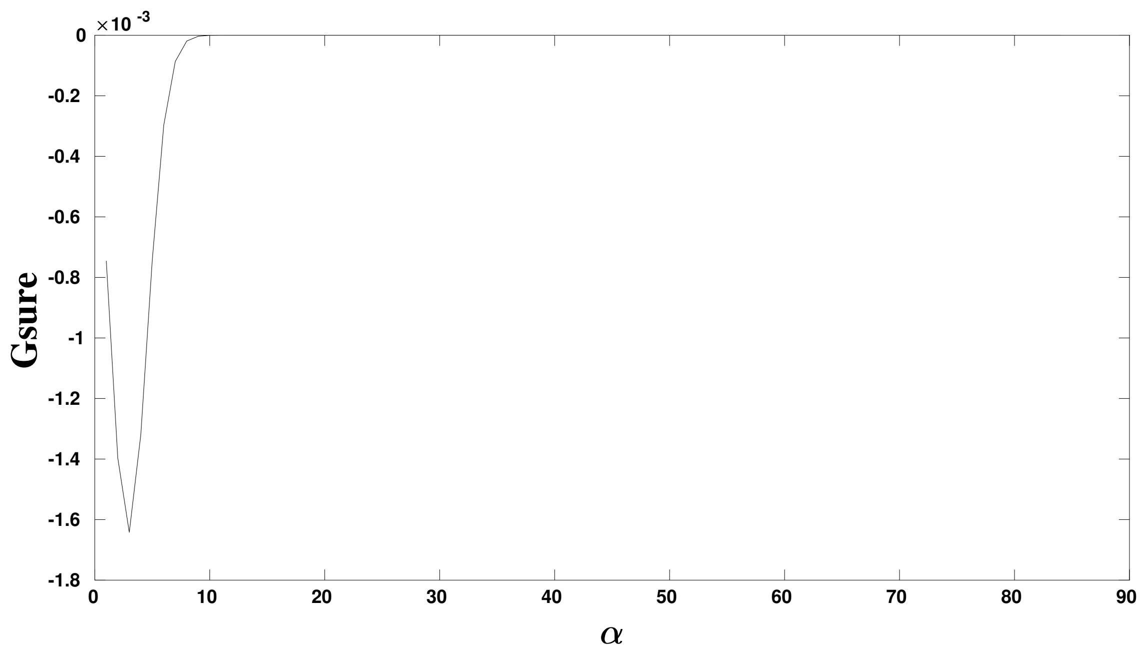

Figure 2: Typical evolution of the GSURE for the parameters n=128, l=0.1 and $\sigma = 0.1$.

9. Acknowledgment

This work was supported by the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

[1] H.W.Engl, M.Hanke and A.Neubauer, Regularization of inverse problems, Kluwer, (1996).

[2] O.Scherzer, M.Grassmair, H.Grossauer, M.Haltmaier and F.Lenzen, Variational methods in imaging, Springer, Verlag, (2008).

[3] S.Mallat, A wavelet tour of signal processing, Elsevier (2009).

[4] L.I.Rudin, S.Osher and E.Fatemi Nonlinear total variation based noise removal algorithms, Physica D: Nonlinear Phenomena, 60 (1992) 259-268 .

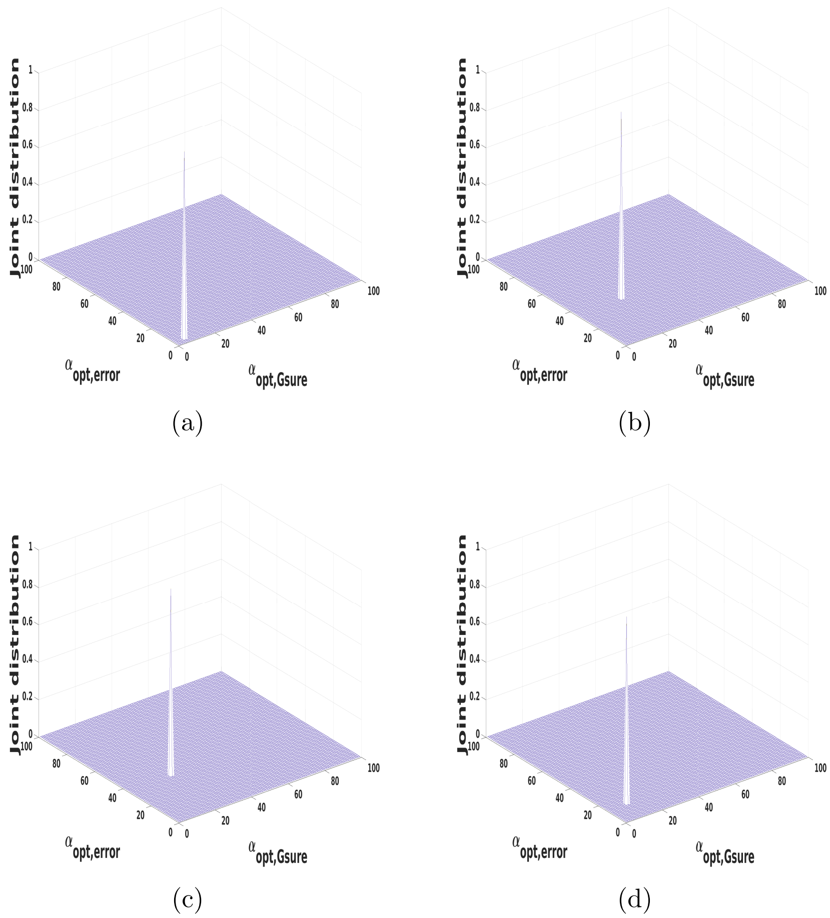

Figure 3: Typical joint distribution of the parameters alpha obtained with GSURE and the minimal l_2 norm for the function f_1 and for the parameters (a) n=32, l=0.05 and $\sigma = 0.01$ (b) n=48, l=0.1 and $\sigma = 0.05$ (c) n=48, l=0.05 and $\sigma = 0.05$ (d) n=32, l=0.1 and $\sigma = 0.05$.

[5] A.Tikhonov and V.Arsenin, Solution, of Ill-posed problems, Winston, Washington, DC, (1977).

[6] L.Lo Gerfo, L.Rosasco, F.Odone, E.De Vito and A.Verri, Spectral algorithms for supervised learning, Neural Computation, 20 (2008) 7.

[7] G.Blanchard, N.Mucke Optimal rates for regularization of statistical

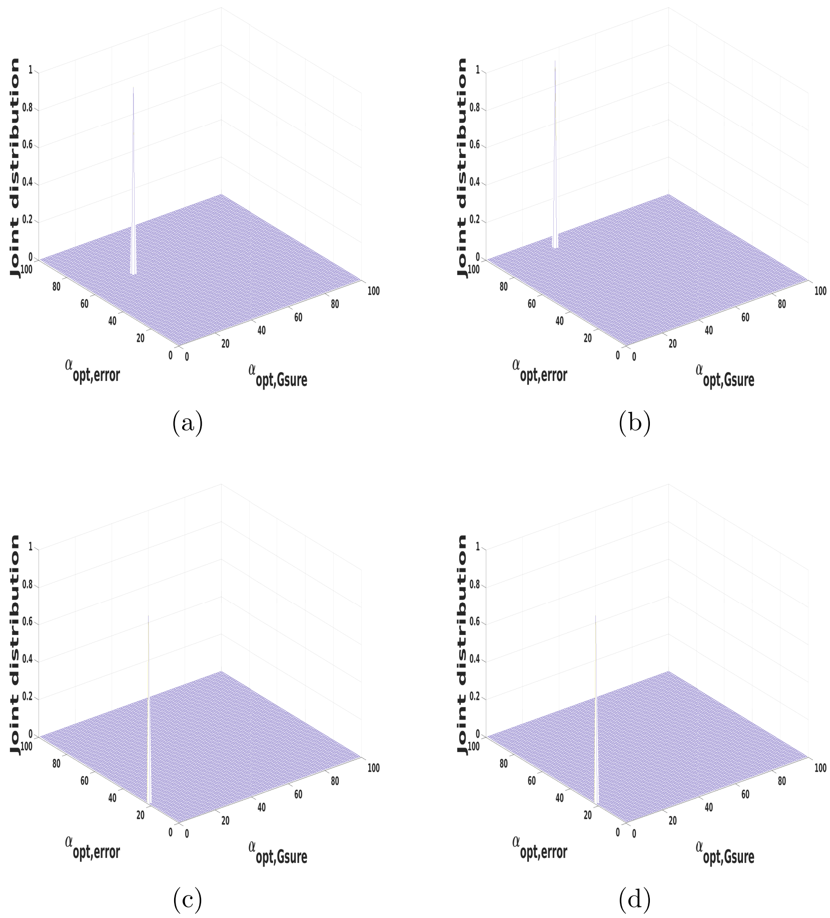

Figure 4: Typical joint distribution of the parameters α obtained with GSURE and the minimal l_2 norm for the function f_1 and for the parameters (a) n=128, l=0.02 and $\sigma = 0.1$ (b) n=128, l=0.01 and $\sigma = 0.1$ (c) n=192, l=0.02 and $\sigma = 0.1$ (d) n=192, l=0.01 and $\sigma = 0.1$.

inverse learning problems, Foundations of Computational Mathematics., 18 (2016) 971-1013 .

[8] F.Dai, H.Zhu, F.H.Wen Two nonparametric approaches to mean absolute deviation portfolio selection problem, Journal of Industrial Management Optimization (2019) doi:10.3934/jimo.2019054.

[9] F.Bauer and M.A.Lukas, Comparing parameter choice methods ofr

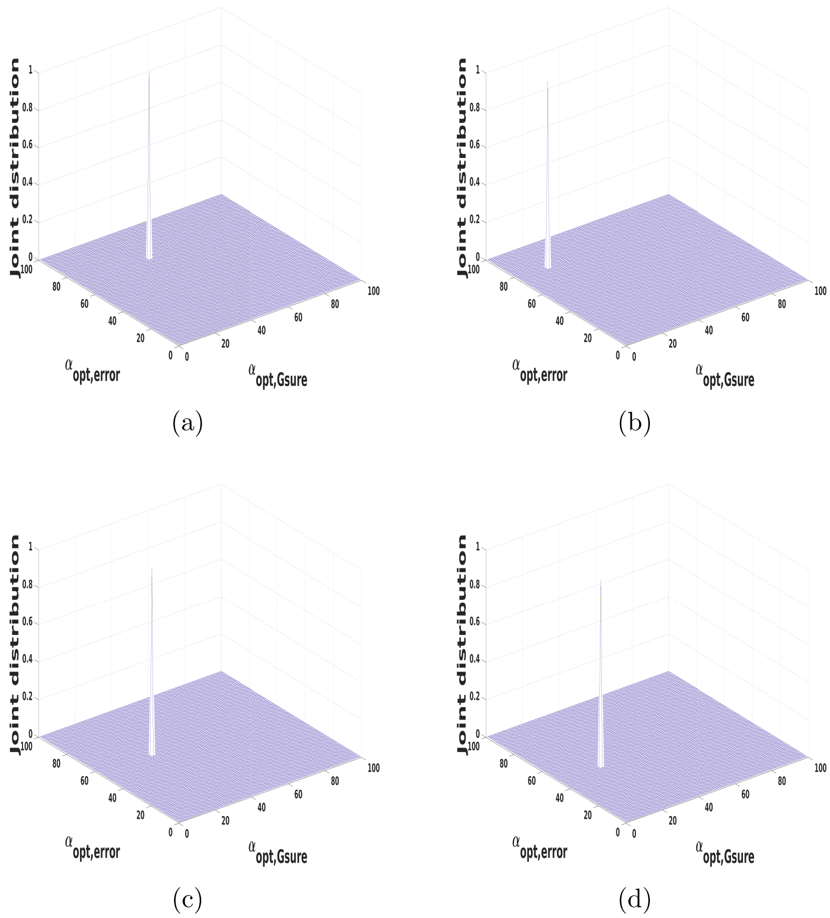

Figure 5: Typical joint distribution of the parameters α obtained with GSURE and the minimal l_2 norm for the function f_2 and for the parameters (a) n=128, l=0.02 and $\sigma = 0.1$ (b) n=128, l=0.01 and $\sigma = 0.1$ (c) n=192, l=0.02 and $\sigma = 0.1$ (d) n=192, l=0.01 and $\sigma = 0.1$.

regularization of ill-posed problems. Mathematics and Computers in Simulation 81 (2011) 1795-1841.

[10] P.C.Hansen Analysis of Discrete Ill-posed Problems by means of the L-curve. SIAM Review, 34 (1992) 561-580.

[11] P.Mathe and S.V.Perevezev, Discretization strategy for ill-posed problems in variable Hilbert scales, Inverse Problems, 19 (2003) 1263-1277.

[12] P.Mathe and S.V.Pereverzev, Regularization of some linear ill-posed problems with discretized random noisy data, Mathematics of Computation, 75 (2006) 1913-1929.

[13] F.Bauer and T.Hohage, A Lepskij-type stopping rule for regularized Newton methods. Inverse Problems, 21 (2005) 1975-1991.

[14] C.M.Stein, Estimation of the Mean of a multivariate normal distribution, The Annals of Statistics, 9 (1981) 1135-1151.

[15] F.Luisier, T.Blu and M.Unser Image denoising in mixed Poisson-Gaussian noise, IEEE Transactions on Image Processing, 20 (2011) 696-708.

[16] J.C.Pesquet, A.Benazza-Benyahia and C.Chaux, A SURE approach for digital signal/image deconvolution problems, IEEE Transactions on Image Processing, 57 (2009) 4616-4632.

[17] S.Ramani, T.Blu and M.Unser, Monte-Carlo Sure: a black box optimization of regularization parameters for general denoising algorithm, IEEE Transactions on Image Processing, 17 (2008) 1540-54.

[18] S.Ramani, Z.Liu, J.Rosen, J.F.Nielsen and J.A.Fessler, Regularization parameter selection for nonlinear iterative image restoration and MRI reconstruction using GCV and Sure based methods, IEEE Transactions on Image Processing, 21 (2012) 3659-72.

[19] Y.Eldar, Generarlized SURE for exponential families applications to regularization, IEEE Transactions on Signal Processing, 57 (2009) 471-481.

[20] C.Vonesch, S.Ramani and M.Unser, Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, ICIP, (2008) .

[21] C.A.Deledalle, S.Vaiter, G.Peyre, J.Fadili and C.Dossal, Unbiased risk estimation for sparse analysis regularization, IEEE International Conference on Image Processing, (2012) 3053-3056.

[22] C.A.Deledalle, S.Vaiter, J.Fadili and G.Peyre, Stein Unbiased Gradient estimator of the Risk (SUGAR) for multiple parameter selection, SIAM Journal on Imaging Science, 7 (2014) 2448-2487.

[23] C.A.Deledalle, Estimation of Kullback-Leibler losses for noisy recovery problems with the exponential family, arXiv preprint 1512.80191.

[24] R.Giryes, M.Elad and Y.C.Eldar, The projected GSURE for automatic parameter tuning in iterative shrinkage methods, Applied and Computational Harmonic Analysis, 30 (2011) 407-422.

[25] F.Lucka, K.Proksch , C.Brune, N.Bissantz , M.Burger, H.Dette, F.Wubbeling, Risk estimators for choosing regularization parameters in ill-posed problems-Properties and Limitations, arXiv:1701.04970v1, 2017.

[26] N.Bissantz, T.Hohage T and A.Munk, Consistency and rates of convergence of nnlinear Tikhonov regularization with random noise, Inverse Problems, 20 (2004) 1773-1789.

[27] N.Bissantz, T.Hohage, A.Munk and F.Ruymgaart, Convergence rates of general regularization mehthods for statistical inverse problems and applications, SIAM J.Numer. Anal., 45 (2007) 2610-2636.

[28] B.Carl and I.Stephani, Entropy, compactness and approximation of operators, Cambridge, University Press, Cambridge, 1990.

[29] K.Atkinson, *Convergence rates for approximate eigenvalues of compact integral operators*, SIAM J.Numer.Anal., **12** (1975) 213-222.

[30] J.E.Osborn, Spectral Approximation for Compact Operators, Mathematics of Computation, 29 (1975) 712-723.

[31] P.M.Anselone, Collectively compact operator approximation theory and applications to integral equations, Prentice-Hall (1971).

[32] R.Dautray and J.L.Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Springer Verlag, Berlin (2000).

[33] T.Blu and F.Luisier, The SURE-LET approach to image denoising, IEEE Transactions on Image Processing, 16 (2007) 2778-2786.

[34] S.Ramani, J.Rosen, Z.Liu and J.A.Fessler, Iterative weighted risk estimation for nonlinear image restoration with analysis priors, Proc.SPIE Elec.Img., 8926 (2012) 82960N1-12.

[35] M.Nair, S.V.Pereverzev and U.Tautenhahn, Regularization in Hilbert scales under general smoothing conditions, SIAM Journal on Numerical Analysis, 21 (2005) 1851-1869.

[36] U.Tautenhahn, On a general regularization scheme for nonlinear ill-posed problems, Inverse Problems, 13 (1997) 1427-1437.

[37] T.Hohage, Regularization of exponentially ill-posed problems, Numer.Funct.Anal.Optim., 21 (2000) 439-464.

[38] P.Mathe and S.V.Perevezev, Moduli of continuity for operator valued functions, Numer.Funct.Anal.Optim., 23 (2002) 623-631.

[39] P.Mathe and S.V.Peveverzev, Geometry of linear ill-posed problems in variable Hilbert scales, Mathematics of Computation, 19 (2003) 789-803.

[40] F.Lucka, K.Proksch, C.Brune, N.Bissantz, M.Burger, H.Dette, F.Wubbeling Risk estimators for choosing regularization parameters in ill-posed problems-properties and limitations, Inverse Problems and Imaging., 12 (2018) 1121-1155.