Convergence of regularization methods with filter functions for a regularization parameter chosen with GSURE

B.Sixou

1 CREATIS, INSALyon; Inserm, U1044; CNRS UMR 5220; Université Lyon 1;
Université de Lyon; 69621, Lyon, France
E-mail: bruno.sixou@insa-lyon.fr

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 but the efficiency increases with the smoothness of the solution.

1. Introduction

In this article, we consider the numerical solution of a linear inverse problem written as $y = Ax$ where $A : X \rightarrow Y$ is an operator belonging to the set $\mathcal{L}(X,Y)$ of linear bounded operators mapping the two infinite dimensional separable Hilbert spaces $X$ and $Y$. We assume that after discretization, the inverse problem is of the form:

$$y^\delta = B_n f + \epsilon$$

where $f \in \mathbb{R}^n$, $y^\delta \in \mathbb{R}^n$ denotes the discrete noisy data, $B_n$ is the discrete approximation of the operator $A$ and $\epsilon$ consists of independent and identically distributed (i.i.d) Gaussian errors with variance equal to $\sigma^2$, $\epsilon \sim \mathcal{N}(0,\sigma^2 I_n)$. In order to obtain a stable solution, it is usual to consider a Tikhonov regularization functional with a regularization parameter $\alpha$:

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

The regularization parameter $\alpha$ has to be chosen carefully[1]. Several rules of choice of the regularization parameters have been investigated in the literature like the L-curve criteria [2] or some discrepancy principles[3]. Rules based on the risk estimation with the Stein Unbiased Risk Estimator (SURE) have been investigated [4, 5, 6, 7]. 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 [8]. The case of denoising was extended to more general linear operators in Vonesch et al.[9]. For general linear inverse problems, some authors have considered a generalized version GSURE where the risk is measured in the space of the unknown[10, 11, 12].
Yet, there are 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 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. 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 is also taken into account.

The outline of the paper is the following. In the first section, we detail 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 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 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 estimates.

2. The SURE principle and its generalizations

In the following, we denote \( f_\alpha(y^\delta) \) the reconstructed solution for the regularization parameter \( \alpha \) and \( f \) the true solution. In inverse problems and especially reconstruction problems, the mean squared error (MSE)

\[
MSE(\alpha) = \frac{\|f - f_\alpha(y^\delta)\|^2}{n}
\]

is a very usual criteria to estimate the quality of the solution of the inverse problem. The Stein lemma is the basis to obtain estimates of this mean square error [4, 13] 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 \( \epsilon \) of variance \( \sigma^2 \), so that \( y^\delta = f + \epsilon \). 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 [4, 5, 6, 14]. It is given by:

\[
SURE(\alpha) = \left\| y^\delta - f_\alpha(y^\delta) \right\|^2/n - \sigma^2 + 2\sigma^2 n^{-1} Tr(J_{f_\alpha}(y^\delta))
\]

where \( J_{f_\alpha}(y^\delta) = \nabla y B_n f_\alpha(y^\delta) \) is the Jacobian matrix of the reconstructed solution. A generalized SURE has been proposed for exponential families [8]. In [14] a weighted mean square error (WMSE) measure has been proposed to test the accuracy of the reconstruction:

\[
WMSE(\alpha) = n^{-1} \left\| y^\delta - B_n f_\alpha(y^\delta) \right\|_W^2
\]

where \( W \) is a positive definite symmetric weighting matrix. The matrix \( W \) is chosen to counterbalance the effects of the direct operator \( B_n \). The operator \( W \) can be written \( W = B^{-1}_n B_n \), where \( B_n \) is an approximate inverse of \( B_n \) depending on the regularization parameter \( \alpha \). When \( B_n \) has full rank, one can chose \( B_n^{-1} = (B_n^* B_n)^{-1} \) and when \( B_n \) is rank deficient, one can set \( B_n^{-1} = B_n^* (B_n B_n^*)^+ \), where \( M^+ \) denotes the pseudo-inverse of \( M \). Under the assumptions of the Stein lemma, the random variable Generalized Sure \( GSURE(\alpha) \) is an unbiased estimator of \( WMSE(\alpha) \) [10, 11, 12]:

\[
GSURE(\alpha) = n^{-1} \left\| B_n^{-1} (y^\delta - B_n f_\alpha(y^\delta)) \right\|^2 - \sigma^2 n^{-1} tr(B_n^{-1} B_n)
+ 2\sigma^2 n^{-1} tr(B_n^{-1} B_n J_{f_\alpha}(y^\delta))
\]

with \( f_\alpha(y^\delta) = B_n y^\delta \). We show in the following that choosing the regularization parameter \( \alpha \) with the minimum of \( GSURE(\alpha) \) given by Eq.6 leads to a convergent regularization method.
3. 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 and the assumptions on the smoothness of the ground truth. The operator $B^*_n B_n$ is a finite rank self-adjoint operator with closed range. It is possible to consider an orthonormal basis of the range of this operator $(e_i)_{1 \leq i \leq n}$.

$$B^*_n B_n f = \sum_{i=1}^{n} \rho_i <e_i, f> e_i$$

where the $\rho_i$ are the square of the singular values $\sigma_i(B_n)$ of the operator $B_n$. We will assume that $n$ is large and that the discretization errors are very small so that the square of the singular values of $B_n$ satisfies a similar rate of decay than the ones of $A$:

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

where $\eta$ is a strictly positive constant. Under this assumption, we can estimate the function $R(\beta)$ defined as $R(\beta) = \Sigma(\{\rho \geq \beta\})$, where $\Sigma$ is the counting measure:

$$R(\beta) = \Sigma(\{\rho \geq \beta\}) = |\{\rho_j \geq \beta\}| \sim \sum_{j \leq \beta^{-\eta}} 1 \sim \int_{0}^{\beta^{-\eta}} dt \sim \beta^{-\eta}$$

The regularized estimators considered are constructed with regularization methods of the form:

$$f_{\alpha} = \Phi_{\alpha}(A^* A)^{-1} A^* y^\delta$$

with $\Phi_{\alpha} : \mathcal{L}(X,Y) \rightarrow \mathcal{L}(X,Y)$ is a mapping between linear operators parametrized by $\alpha$. With the discretization of the operator $A$, the approximate solution is calculated as:

$$f_{\alpha} = \Phi_{\alpha}(B^*_n B_n)^{-1} B^*_n y^\delta.$$  

The regularized inverse will be defined as $B_{\text{inv}} = \Phi_{\alpha}(B^*_n B_n)^{-1}$. In order to define precisely $\Phi_{\alpha}(B^*_n B_n)$, we will use the spectral family of the operator $B^*_n B_n$[19]. We decompose the space $H = \mathbb{R}^n$ into a direct sum of subspaces $V_k$ in which $B^*_n B_n$ is reduced to the multiplication by the eigenvalue $\rho_k$ of $B^*_n B_n$, $\mathbb{R}^n = \bigoplus V_k$. We denote $P_k$ the orthogonal projection operator onto $V_k$ and we introduce, for all $\rho \in \mathbb{R}$ the space $G_\rho = \bigoplus_{\rho_k \leq \rho} V_k$. Let $E_\rho$ the orthogonal projection onto $G_\rho$. The discontinuities of the function $\rho \rightarrow E_\rho$ are the eigenvalues $\rho_k$. The derivative of $E_\rho$ can be identified as a measure $dE_\rho$ given by:

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

where $\otimes$ is the tensorial product. Let $\phi_\alpha(t)$ a piecewise continuous function, approximating the function $t \rightarrow 1/t$, and which is parametrized by a regularization parameter $\alpha$, the operator $\Phi_{\alpha}(B^* B)$ is defined by:

$$\Phi_{\alpha}(B^* B) = \int_{0}^{a} \phi_\alpha(\rho)dE_\rho$$

where $a$ a constant such $\|B^* B\| \leq a$. 

Several regularization methods can be described with these type of filter functions. For $(1 - t\alpha(t)) \Lambda(t)$ there exist some positive constant $C$ such that:

$$\lim_{t \to 0} \frac{\phi_{\alpha}(t)}{t} = 1/t, \sup_{t \in \sigma(A^*A)} |t^{1/2} \phi_{\alpha}(t)| \leq C \sqrt{t} 0 < \alpha < \|A^*A\|$$

(14)

$$\sup_{t \in \sigma(A^*A)} |1 - t\phi_{\alpha}(t)| \leq C_2 \ 0 < \alpha < \|A^*A\|, 0 \leq \phi_{\alpha}(t) \leq 1/\alpha \ \forall t \in \sigma(A^*A)$$

(15)

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 $t \geq \alpha$, and $\phi_{\alpha}(t) = 0$ for $t \leq \alpha$.

The convergence rate for the reconstruction methods is determined by some a priori assumption on the exact solution $f$. Following [17], we will measure the smoothness of the function $f$ relative to the smoothing properties of $A$ with a source condition. We assume there exists $w \in X$, and $R > 0$ such that:

$$f = \Lambda(A^*A)w \ \|w\| \leq R$$

(16)

with $\Lambda : \sigma(A^*A) \to [0, \infty]$ a continuous, strictly increasing index function with $\Lambda(0) = 0$. The error estimates are different under different assumptions on the index function. A common assumption is $\Lambda(t) = t^\nu$ with $\nu > 0$ for finitely smoothing operators and mildly ill-posed problems. 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 $c_1 A_2 \leq A_1 \leq c_2 A_2$. We also assume that:

$$\sup_{t \geq 0} |(1 - t\phi_{\alpha}(t)) \Lambda(t)| \asymp \Lambda(\alpha) \ 0 < \alpha < \|A^*A\|$$

(17)

for the index functions associated to the smoothness class. The assumption of Eq.15 can be easily verified for the former regularization methods with $\Lambda(t) = \lambda^\beta$, $0 < \beta < 1$. In the following, we will focus on the Tikhonov regularization method. For this smoothness class and exact data $y$, using Eq.15, we see that the error is bounded by [18]:

$$\|\Phi_{\alpha}(A^*A)A^*y - f\| \leq C \sup_{t \in \sigma(A^*A)} |(\phi_{\alpha}(t)t - 1)\Lambda(t)| \leq CA(\alpha)$$

(18)

4. Estimate of the risk GSURE

In this section, we present some estimation of GSURE for Tikhonov regularization. Similar estimations can be derived for the filter based regularization presented above. Let $e_i$ be an eigenvector for the eigenvalue $\rho_i$, with Eq.13, we have:

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

(19)

We start from Eq.6 to estimate GSURE:

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

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

(20)

(21)

Using the spectral projectors, and for $\phi_{\alpha}$ the filter function for the Tikhonov regularization, we get:

$$\|\Phi_{\alpha}(B_n^*B_n)B_n^*e_i\|^2 = \int_0^\infty \phi_{\alpha}^2(\rho)d\rho \|E_{\rho}(e_i)\|^2 = \int_0^\infty \frac{\rho}{(\rho + \alpha)^2} d\|E_{\rho}(e_i)\|^2$$

(22)
We will split this integral in Eq.22 in two terms. For $0 < \rho < \alpha$, we have $1/4 \leq \frac{1}{(1+\rho/\alpha)^2} \leq 1$ and thus:

$$I_1 = \sum_i \int_{\rho<\alpha} \frac{\rho}{(\rho + \alpha)^2} d\|E_\rho(e_i)\|^2 = \frac{1}{\alpha^2} \sum_i \int_{\rho<\alpha} \frac{\rho}{(1+\rho/\alpha)^2} d\|E_\rho(e_i)\|^2 \approx \frac{1}{\alpha^2} \sum_i \int_{\rho<\alpha} \rho d\|E_\rho(e_i)\|^2 \approx \frac{1}{\alpha^2} \sum_i \int_{\rho<\alpha} \rho \delta(\rho - \rho_i) \tag{23}$$

We can reformulate it with the Stieltjes integral:

$$I_1 \approx -\frac{1}{\alpha^2} \int_0^{\rho<\alpha} \beta dR(\beta) \tag{24}$$

Let us assume that $R$ is smooth with $R(\beta) \sim \beta^{-\eta}$, we obtain: $I_1 \approx \frac{1}{\alpha} (\alpha)^{1-\eta} \times \alpha^{-1-\eta}$. The same result can be obtained 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:

$$I_2 = \sum_i \int_{\rho>\alpha} \frac{\rho}{(\rho + \alpha)^2} d\|E_\rho(e_i)\|^2 = \sum_i \int_{\rho>\alpha} \frac{1}{\rho(1+\alpha/\rho)^2} d\|E_\rho(e_i)\|^2 \tag{25}$$

Similarly, the term $I_2$ can be written as $I_2 \approx -\int_{\rho>\alpha} \frac{1}{\rho} dR(\beta) \times \alpha^{-1-\eta}$ we obtain thus

$$Tr(B_{inv}B_{inv}^*) \approx \alpha^{-1-\eta} \tag{26}$$

With the spectral projectors, we obtain similarly:

$$I = Tr(B_{inv}B_{inv}^*B_{inv}) = Tr[(B_nB_n^*)^2 \phi_\alpha(B_n^*B_n)^3] \tag{27}$$

$$= \sum_{1 \leq i \leq n} \|\phi_\alpha^{\frac{3}{2}}(B_n^*B_n)B_n^*B_n e_i\|^2 = \sum_i \int_0^\alpha \rho^2 \phi_\alpha^2(\rho) d\|E_\rho(e_i)\|^2 \tag{28}$$

We split the integral in two parts:

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

$$1_2 = \sum_i \int_\alpha^\infty \rho^2 \phi_\alpha^2(\rho) d\|E_\rho(e_i)\|^2 \approx \sum_i \int_\alpha^\infty \frac{1}{\rho^3} d\|E_\rho(e_i)\|^2 \approx -\int_\alpha^\infty \frac{1}{\rho} dR(\rho) \times \alpha^{-1-\eta} \tag{29}$$

$$Tr(B_{inv}B_{inv}^*B_{inv}) \approx \alpha^{-1-\eta} \tag{30}$$

We set $f_{\alpha} = \Phi_\alpha(B_n^*B_n)B_n^*B_nf$ and $f_{\alpha} = \Phi_\alpha(B_n^*B_n)B_n^*B_ne$, with $f_\alpha = f_{\alpha}^1 + f_{\alpha}^2$. Using the spectral projectors, it can be demonstrated that the average value $E[\|f_{\alpha}^1 - B_{inv}B_nf_{\alpha}^1\|]$ is negligible with respect to $\|f_{\alpha}^1 - B_{inv}B_nf_{\alpha}^1\|$. Using concentration inequalities, we consider we can neglect $\|f_{\alpha}^2 - B_{inv}B_n f_{\alpha}^2\|$:

$$\|f_\alpha - B_{inv}B_nf_\alpha\| \approx \|f_{\alpha}^1 - B_{inv}B_nf_{\alpha}^1\| = \|\phi_\alpha(B_n^*B_n)B_n^*B_n f\| \approx \Lambda(\alpha) \tag{31}$$

Using Eq.27,31,33 GSURE can be estimated as:

$$GSURE(\alpha) \approx \frac{1}{n}(\alpha^{2\mu} + C_1 \sigma^2 \alpha^{-1-\eta}) \tag{33}$$

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

5
5. Discussion and conclusion

We have estimated 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. The convergence of the Tikhonov regularization method is obtained for $\alpha \to 0$, and $\frac{\sigma^2}{\alpha} \to 0[3]$. The choice $\alpha \sim \frac{\sigma^2}{1+\eta+2\mu}$ obtained with the minimization of GSURE ensures thus the convergence of the Tikhonov regularization for a regularization parameter chosen with this estimator. Moreover, it has been shown that for index functions of the type $\Lambda(t) = t^\mu$, the best possible accuracy or convergence rate is of order $\frac{\sigma^2 \mu}{1+2\mu}$ [18]. Since $\frac{2\mu}{1+\eta+2\mu} < \frac{2\mu}{1+2\mu}$, we can conclude that a regularization parameter chosen with the GSURE is not optimal with respect to the convergence rate as the noise level tends towards 0. As $\eta$ increases, the ill-posedness increases and the efficiency of the method of choice of the parameter decreases. If the condition of $B_n$ increases, smaller singular values are added and it is expected that the GSURE is less efficient.

References

[1] F.Bauer and M.A.Lukas, Comparing parameter choice methods for regularization of ill-posed problems. Mathematics and Computers in Simulation 81 (2011) 1795-1841.
[2] P.C.Hansen Analysis of Discrete Ill-posed Problems by means of the L-curve. SIAM Review, 34 (1992) 561-580.
[3] O.Scherzer,M.Grassmair, H.Grossauer, M.Haltmaier and F.Lenzen, Variational Methods in Imaging, Springer Verlag, New York (2008).
[4] C.M.Stein, Estimation of the Mean of a multivariate normal distribution, The Annals of Statistics, 9 (1981) 1135-1151.
[5] F.Luisier, T.Blu and M.Unser Image denoising in mixed Poisson-Gaussian noise, IEEE Transactions on Image Processing, 20 (2011) 696-708.
[6] 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.
[7] 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.
[8] Y.Eldar, Generalized SURE for exponential families applications to regularization, IEEE Transactions on Signal Processing, 57 (2009) 471-481.
[9] C.Vonesch, S.Ramani and M.Unser, Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, ICIP, (2008).
[10] 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) 2484-2487.
[11] C.A.Deledalle, Estimation of Kullback-Leibler losses for noisy recovery problems with the exponential family, arXiv preprint 1512.80191.
[12] 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.
[13] T.Blu and F.Luisier, The SURE-LET approach to image denoising, IEEE Transactions on Image Processing, 16 (2007) 2778-2786.
[14] 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.
[15] 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.
[16] U.Tautenhahn, On a general regularization scheme for nonlinear ill-posed problems, Inverse Problems, 13 (1997) 1427-1437.
[17] N.Bissantz, T.Hohage T and A.Munk, Consistency and rates of convergence of nonlinear Tikhonov regularization with random noise, Inverse Problems, 20 (2004) 1773-1789.
[18] 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.
[19] R.Dautray and J.L.Lions, Mathematical Analysis and Numerical Methods for Science and Technology, Springer Verlag, Berlin (2000).