Local Behavior of Sparse Analysis Regularization: Applications to Risk Estimation

Samuel Vaiter, Charles Deledalle, Gabriel Peyré, Charles H Dossal, Jalal M. Fadili

Local Behavior of Sparse Analysis Regularization: Applications to Risk Estimation

Samuel Vaiter, Charles-Alban Deledalle, Gabriel Peyré

CEREMADE, CNRS, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France

Charles Dossal

IMB, Université Bordeaux 1, 351, Cours de la libération, 33405 Talence Cedex, France

Jalal Fadili

GREYC, CNRS-ENSICAEN-Université de Caen, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France

Abstract

This paper studies the recovery of an unknown signal x_0 from low dimensional noisy observations $y = \Phi x_0 + w$, where Φ is an ill-posed linear operator and w accounts for some noise. We focus our attention to sparse analysis regularization. The recovery is performed by minimizing the sum of a quadratic data fidelity term and the ℓ^1-norm of the correlations between the sought after signal and atoms in a given (generally overcomplete) dictionary. The ℓ^1 prior is weighted by a regularization parameter $\lambda > 0$ that accounts for the noise level. In this paper, we prove that minimizers of this problem are piecewise-affine functions of the observations y and the regularization parameter λ. As a byproduct, we exploit these properties to get an objectively guided choice of λ. More precisely, we propose an extension of the Generalized Stein Unbiased Risk Estimator (GSURE) and show that it is an unbiased estimator of an appropriately defined risk. This encompasses special cases such as the prediction risk, the projection risk and the estimation risk. We also discuss implementation issues and propose fast algorithms. We apply these risk estimators to the special case of sparse analysis regularization. We finally illustrate the applicability of our framework on several imaging problems.

Key words: sparsity, analysis regularization, inverse problems, ℓ^1 minimization, local variation, degrees of freedom, SURE, GSURE, unbiased

Email addresses: samuel.vaiter@ceremade.dauphine.fr (Samuel Vaiter), deledalle@ceremade.dauphine.fr (Charles-Alban Deledalle), gabriel.peyre@ceremade.dauphine.fr (Gabriel Peyré), charles.dossal@math.u-bordeaux1.fr (Charles Dossal), Jalal.Fadili@greyc.ensicaen.fr (Jalal Fadili)

risk estimation.

1. Introduction

1.1. Linear Inverse Problems

In many applications, the goal is to recover an unknown signal $x_0 \in \mathbb{R}^N$ from noisy and degraded observations $y \in \mathbb{R}^Q$. The forward observation model

$$y = \Phi x_0 + w, \tag{1}$$

assumes that the degradation is linear. The noise is a deterministic vector $w \in \mathbb{R}^Q$ for the contributions detailed in Section (2.1), and a zero-mean white Gaussian noise $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$ for the contributions of Sections 2.2, 2.3 and 2.4. The mapping $\Phi : \mathbb{R}^N \to \mathbb{R}^Q$ is a bounded linear operator which is typically ill-behaved since it models an acquisition process that entails loss of information so that $Q \leqslant N$. In image processing, typical cases covered by the above degradation model are entry-wise masking (inpainting), convolution (acquisition blur), Radon transform (tomography) or a random sensing matrix (compressed sensing).

Linear inverse problems are among the most active fields in signal and image processing [1]. In order to regularize them and reduce the space of candidate solutions, one has to incorporate some prior knowledge on the typical structure of the original signal or image x_0. This prior information accounts for the smoothness of the solution and can range from uniform smoothness assumption to more complex geometrical priors.

1.2. Variational Regularizations

Variational analysis gives a framework to inverse linear problems such as (1). The general method reads

$$x^\star(y) \in \operatorname*{argmin}_{x \in \mathbb{R}^N} F(x, y) + \lambda R(x). \tag{2}$$

where F is a so-called data fidelity term, R a regularization of the recovered signal and $\lambda > 0$ a regularization parameter. This term allows one to balance the impact of the regularization in the minimization. In this paper, we consider a least square error term which reads

$$F(x, y) = \frac{1}{2}\|y - \Phi x\|_2^2, \tag{3}$$

and corresponds in a bayesian framework to a Gaussianity assumption on the noise w. Tikhonov regularization makes use of a quadratic prior $R(x) = \langle x, Kx \rangle$, where K is a symmetric positive definite kernel. This typically enforces some kind of uniform smoothness of the recovered vector. To capture the complexity of image structures, non-quadratic priors are required, among which sparse regularizations using the ℓ^1 is the most popular choice.

1.3. Sparse Analysis Regularization

We call a dictionary $D = (d_i)_{i=1}^P$ a collection of P atoms $d_i \in \mathbb{R}^N$. This collection may be redundant in $\mathbb{R}^Q$, whose span may be $\mathbb{R}^N$ or only a subset of it. It can also be viewed as a linear mapping from $\mathbb{R}^P$ to $\mathbb{R}^N$ which is used to *synthesize* a signal $x \in \mathrm{Span}(D) \subseteq \mathbb{R}^N$ as $x = D\alpha = \sum_{i=1}^P \alpha_i d_i$, where α is not uniquely defined if D is a redundant dictionary.

The analysis regularization with respect to a dictionary D corresponds to using $R = R_A$ in (2) where

$$R_A(x) = \|D^* x\|_1. \tag{4}$$

This leads us to the following minimization problem which is the focus of this paper

$$x^\star(y) \in \min_{x \in \mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \|D^* x\|_1. \tag{$\mathcal{P}_\lambda(y)$}$$

Note that the set of (global) minimizers of $\mathcal{P}_\lambda(y)$ is nonempty and compact if, and only if,

$$\mathrm{Ker}\, \Phi \cap \mathrm{Ker}\, D^* = \{0\}, \tag{H_0}$$

since the objective function is proper, continuous and convex (see for instance [2]). All throughout this paper, we suppose that this condition holds.

The most popular analysis sparse regularization is the total variation, which was first introduced for denoising in [3]. It corresponds to using an operator D^* which is a finite differences approximation of the gradient. The corresponding prior R_A favors piecewise constant signals and images. A review of total variation regularization can be found in [4]. The theoretical properties of total variation for denoising has been extensively studied. A distinctive feature of this regularization is that it tends to produces a staircasing effect, where discontinuities not present in the original data might be created by the regularization. This effect has been studied by Nikolova in [5] in 2-D.

When $D = \mathrm{Id}$, the sparse prior R_A is coined synthesis regularization. The corresponding regularization problem is either referred to as the LASSO problem in the statistics community [6] and basis-pursuit denoising in the signal community [7]. Despite synthesis and analysis priors both making use of the ℓ^1 norm, their behaviors differ greatly when D is not orthogonal, see for instance [8]. While the theoritical performance of synthesis regularization has been extensively studied, only a few papers have targeted the analysis case, see in particular [9, 10].

1.4. Local Variations

Local variations and sensitivity/perturbation analysis of problems in the form (2) is an important topic in optimization and optimal control. Comprehensive monograph on the subject are [11, 12]. In this paper, we focus on the variations with respect to the regularization parameter λ and the observations y, i.e we study the set-valued mapping $(\lambda, y) \mapsto \mathcal{M}_\lambda(y)$ where $\mathcal{M}_\lambda(y)$ is the set of minimizers of (2). We also restrict our attention to first-order properties of

this mapping, although second-order properties might be of interest as well (see e.g. [11, 13]).

In the synthesis case ($D = \mathrm{Id}$), the works of [14, 15] prove that the mapping $\lambda \mapsto x^\star(y)$ is piecewise affine for a fixed y. This enables the computation of the set of solutions for all λ using an homoropy algorithm. This result is extended to the underdetermined case in [16]. The work of [17] proposes a homotopy algorithm in the overdetermined case for sparse analysis regularization. We come back to this latter work in Section 3.

1.5. Risk Estimation

This paper is also focussed on unbiased estimation of the ℓ^2-risk of recovering a vector $x_0 \in \mathbb{R}^N$ from (1) by solving (2), under the assumption that w is a Gaussian white noise. These unbiased estimates depend solely on y, without prior knowledge of x_0. This can prove very useful as a basis for automatic ways to choose the parameters of the reconstruction algorithm, e.g. λ in (2).

Degrees of freedom (DOF) is a familiar phrase in statistics. More generally, degrees of freedom is often used to quantify the complexity of a statistical modeling procedure. However, there is no exact correspondence between the DOF and the number of parameters in the model. The DOF plays an important role in model validation and selection. From the seminal definition of Efron [18], the degrees of freedom is given by

$$df(x^\star) = \sum_{i=1}^{Q} \frac{\mathrm{cov}(y_i, (\Phi x^\star(y))_i)}{\sigma^2} \ .$$

Many model selection criteria involve the DOF, e.g. AIC (Akaike information criterion [19]), BIC (Bayesian information criterion [20]) or GCV (generalized cross-validation [21]). It allows us to estimate the risk in reconstructing Φx_0, i.e. the *prediction* risk $R = \mathbb{E}_w(\|\Phi x^\star(y) - \Phi x_0\|^2)$. Indeed, Mallows' C_p statistic

$$C_p = \|y - \Phi x^\star(y)\|^2 - Q\sigma^2 + 2\sigma^2 df(x^\star)$$

is an unbiased estimate of $R = \mathbb{E}_w(C_p)$. Since the DOF is usually unknown, C_p cannot be used directly. Instead, an unbiased estimate of the DOF can be used to unbiasedly estimate the risk through a modified C_p, e.g. the Stein Unbiased Risk Estimator (SURE) given by [22]:

$$\mathrm{SURE} = \|y - \Phi x^\star(y)\|^2 - Q\sigma^2 + 2\sigma^2 \hat{df}(x^\star) \quad \text{with} \quad \hat{df}(x^\star) = \mathrm{tr}\left(\frac{\partial \Phi x^\star(y)}{\partial y}\right) \quad (5)$$

and where $\mathbb{E}_w(\hat{df}(x^\star)) = df(x^\star)$.

1.6. Applications of (G)SURE

Applications of SURE emerged for choosing the smoothing parameters in families of linear estimates [23] such as for model selection, ridge regression, smoothing splines, etc. It has been extensively used in the statistical community

as a competitor to other model selection techniques, e.g. AIC, BIC and GCV. In some setting, it has been shown that it offers better accuracy than GCV and related non-parametric selection techniques [24]. Compared to GCV, the drawback of SURE is that it requires the knowledge of the noise variance σ^2.

After its introduction in the wavelet community with the SURE-Shrink algorithm [25], it has been widely used for various image denoising problems, e.g. in sparse regularization [26, 27, 28] and in non-local filtering [29, 30, 31]. In the context of inverse problems, the minimizers of the prediction risk can sometimes be far away from the minimizers of the *estimation* risk $\mathbb{E}(\|x^\star(y) - x_0\|^2)$ [32]. In [27], the authors proposed an approximation of the estimation risk that relies on a stabilized approximation of the inverse of Φ. In general, either Φ should have full rank or x_0 should belong to $\ker(\Phi)^\perp$ to guarantee the existence of an unbiased estimator of the estimation risk [33].

A generalized SURE (GSURE) has been developed for noise models within the multivariate canonical exponential family [34]. It allows one to estimate the risk on a projected version of $x^\star$. Indeed, in the scenario where Φ is rank-deficient or redundant, the GSURE can at best estimate the *projection* risk $\mathbb{E}(\|\Pi x^\star(y) - \Pi x_0)\|^2)$ where Π is the orthogonal projector on $\ker(\Phi)^\perp$.

1.7. Organization of this Paper

Section 2 details our three contributions. Section 3 draws some connections with relevant previous works. Section 4 illustrates our results using numerical examples. The proofs are deferred to Section 5 awaiting inspection by the interested reader.

2. Contributions

The contributions at the heart of this paper are the following:

(i) **Local affine parameterization:** a solution of $\mathcal{P}_\lambda(y)$ is a piecewise affine function of (y, λ).

(ii) **GSURE:** an unifying framework to compute unbiased estimates of several risks, including the prediction risk, the projection risk and the estimation risk.

(iii) **Sparse Analysis Estimation Risk:** the previous framework is instantiated for sparse analysis-based estimators.

(iv) **Numerical Computation of GSURE:** sparse analysis DOF and GSURE can be approximated by solving a simple linear system.

Each contribution is rigorously described in the following sub-sections.

We start with some notations used in the sequel. The sign vector $\mathrm{sign}(\alpha)$ of $\alpha \in \mathbb{R}^P$ is

$$\forall k \in \{1, \cdots, P\}, \quad \mathrm{sign}(\alpha)_k = \begin{cases} +1 & \text{if} \quad \alpha_k > 0, \\ 0 & \text{if} \quad \alpha_k = 0, \\ -1 & \text{if} \quad \alpha_k < 0. \end{cases}$$

The support of $\alpha \in \mathbb{R}^P$ is

$$\operatorname{supp}(\alpha) = \{i \in \{1, \cdots, P\} \setminus \alpha_i \neq 0\}.$$

For a set I, $|I|$ denotes the cardinal of I. The matrix M_J for J a subset of $\{1, \ldots, P\}$ is the submatrix whose columns are indexed by J. Similarly, the vector s_J is the reduced dimensional vector built upon the components of s indexed by J. The matrix Id is the identity matrix, where the underlying space is implicitly defined from the context. For any matrix M, M^+ is the Moore–Penrose pseudoinverse of M and M^* is the adjoint matrix of M.

2.1. Local Affine Parameterization

In this section, the noise $w \in \mathbb{R}^Q$ is a deterministic vector. Our first contribution gives a local affine parameterization of solutions of $\mathcal{P}_\lambda(y)$.

Recall that D is a dictionary of $\mathbb{R}^{N \times P}$. We define the *D-support* I (resp. *D-cosupport* J) of a vector $x \in \mathbb{R}^N$ as $I = \operatorname{supp}(D^* x)$ (resp. $J = I^c$). Given J a subset of $\{1 \cdots P\}$, the *cospace* $\mathcal{G}_J$ is defined as

$$\mathcal{G}_J = \operatorname{Ker} D_J^*.$$

For some cosuport J, it is important to ensure the invertibility of Φ on $\mathcal{G}_J$. This is achieved by imposing

$$\operatorname{Ker} \Phi \cap \mathcal{G}_J = \{0\}. \tag{H_J}$$

Note that there is always a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds as shown in Lemma 6.

Definition 1. *Let J be a D-cosupport. Suppose that (H_J) holds. We define the operator $\Gamma^{[J]}$ as*

$$\Gamma^{[J]} = U \left(U^* \Phi^* \Phi U\right)^{-1} U^*. \tag{6}$$

where U is a matrix whose columns form a basis of $\mathcal{G}_J$.

The transition space $\mathcal{H}$ defined below corresponds to observations y and scaling parameter λ where the cospace $\mathcal{G}_J$ of the solution of $\mathcal{P}_\lambda(y)$ is not stable with respect to small perturbations of (y, λ).

Definition 2. *The* transition space $\mathcal{H}$ *is defined as*

$$\mathcal{H} = \bigcup_{\substack{J \subset \{1, \cdots, P\} \\ (H_J) \text{ holds}}} \bigcup_{\substack{K \subset J \\ \operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}}} \bigcup_{s_{J^c} \in \{-1,1\}^{|J^c|}} \bigcup_{s_K \in \{-1,1\}^{|K|}} \mathcal{H}_{J,K,s_{J^c},s_K},$$

where

$$\mathcal{H}_{J,K,s_{J^c},s_K} = \left\{(y, \lambda) \in \mathbb{R}^Q \times \mathbb{R} \setminus P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} y = \lambda(\tilde{\Omega}^{[J]} s_{J^c} + D_K \sigma_K)\right\},$$

where $\tilde{\Pi}^{[J]} = \Phi^*(\Phi \Gamma^{[J]} \Phi^* - \mathrm{Id})$, $\tilde{\Omega}^{[J]} = \Phi^* \Phi \Gamma^{[J]} - \mathrm{Id}$ and $P_{\mathcal{G}_{K^c}}$ is the orthogonal projection on $\mathcal{G}_{K^c}$.

The following theorem is our first contribution.

Theorem 1. *Let $(y, \lambda) \notin \mathcal{H}$ and let $x^\star(y)$ a solution of $\mathcal{P}_\lambda(y)$. Let I be the D-support and J the D-cosupport of $x^\star(y)$ and $s = \mathrm{sign}(D^* x^\star(y))$. We suppose that (H_J) holds. We define*

$$\forall \bar{y} \in \mathbb{R}^Q, \forall \bar{\lambda} \in \mathbb{R}, \quad \hat{x}_{\bar{\lambda}}(\bar{y}) = \Gamma^{[J]} \Phi^* \bar{y} - \bar{\lambda} \Gamma^{[J]} D_I s_I.$$

There exists an open neighborhood $\mathcal{B} \subset \mathbb{R}^Q \times \mathbb{R}$ of (y, λ) such that for every $(\bar{y}, \bar{\lambda}) \in \mathcal{B}$, $\hat{x}_{\bar{\lambda}}(\bar{y})$ is a solution of $\mathcal{P}_{\bar{\lambda}}(\bar{y})$.

If $\mathcal{P}_\lambda(y)$ admits a unique solution $x_\lambda(y)$ for each λ, this theorem shows that $\lambda \mapsto x_\lambda(y)$ is a polygonal path in $\mathbb{R}^N$.

2.2. Generalized Stein Unbiased Risk Estimator

For the following contributions, the noise is assumed to be a zero-mean white Gaussian vector $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$. In this section, we consider an arbitrary estimator $x^\star(y)$ such that $\Phi x^\star(y)$ is defined without ambiguity. Using this assumption, we define the quantities

$$\mu^\star(y) = \Phi x^\star(y) \quad \text{and} \quad \mu_0 = \Phi x_0.$$

We define an extension of GSURE that estimates the risk of reconstructing $A\mu_0$ with an arbitrary matrix $A \in \mathbb{R}^{M \times Q}$. We introduce this general definition which allows one to recover the prediction risk (with $A = \mathrm{Id}$), the projection risk when Φ is rank deficient (with $A = \Phi^*(\Phi\Phi^*)^+$) and the estimation risk when Φ has full rank (with $A = (\Phi^*\Phi)^{-1}\Phi^*$).

Definition 3. *Let $A \in \mathbb{R}^{M \times Q}$. We define the Generalized Stein Unbiased Risk Estimate (GSURE) associated to A as*

$$\mathrm{GSURE}^A(y) = \|A(y - \mu^\star(y))\|^2 - \sigma^2 \,\mathrm{tr}(A^*A) + 2\sigma^2 \mathrm{dev}_A(y) \ ,$$

where

$$\mathrm{dev}_A(y) = \mathrm{tr}\left(A \frac{\partial \mu^\star(y)}{\partial y} A^* \right) \ .$$

The next theorem is our second contribution, and shows the importance of this estimator.

Theorem 2. *Let $A \in \mathbb{R}^{M \times Q}$. Suppose $y \mapsto \mu^\star(y)$ is weakly differentiable. If $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$, then*

$$\mathbb{E}_w\left(\mathrm{GSURE}^A(y)\right) = \mathbb{E}_w\left(\|A\mu_0 - A\mu^\star(y)\|^2\right) \ .$$

All estimators of the form GSURE^B with B such that $B\Phi = A\Phi$ share the same expectation given by Theorem 2. Hence, there are several ways to estimate the risk in reconstructing $A\mu_0$. For the estimation of the prediction, the projection and the estimation risks, we introduce the following *canonical* estimators (with subscript notations) as direct consequences of Theorem 2:

• GSURE^{Id} provides an unbiased estimate of the prediction risk:

$$\text{GSURE}_{\Phi}(y) = \|y - \mu^\star(y)\|^2 - Q\sigma^2 + 2\sigma^2 \operatorname{tr}\left(\frac{\partial \mu^\star(y)}{\partial y}\right)$$

which coincides with the classical SURE defined in eq. (5).

• When Φ is rank deficient, $\Pi = \Phi^*(\Phi\Phi^*)^+\Phi$ is the orthogonal projector on $\ker(\Phi)^\perp = \operatorname{Im}(\Phi^*)$. Denoting $x_{ML}(y) = \Phi^*(\Phi\Phi^*)^+ y$ the maximum likelihood estimator, $\text{GSURE}^{\Phi^*(\Phi\Phi^*)^+}$ provides an unbiased estimate of the projection risk:

$$\text{GSURE}_{\Pi}(y) = \|x_{ML}(y) - \Pi x^\star(y)\|^2 - \sigma^2 \operatorname{tr}\big((\Phi\Phi^*)^+\big) + 2\sigma^2 \operatorname{tr}\left((\Phi\Phi^*)^+ \frac{\partial \mu^\star(y)}{\partial y}\right).$$

• When Φ has full rank, $y \mapsto x^\star(y)$ is uniquely defined and weakly differentiable. Denoting $x_{ML}(y) = (\Phi^*\Phi)^{-1}\Phi^* y$ the maximum likelihood estimator, $\text{GSURE}^{(\Phi^*\Phi)^{-1}\Phi^*}$ provides an unbiased estimate of the estimation risk given by:

$$\begin{aligned}\text{GSURE}_{\text{Id}}(y) = &\|x_{ML}(y) - x^\star(y)\|^2 - \sigma^2 \operatorname{tr}\left((\Phi^*\Phi)^{-1}\right) \\ &+ 2\sigma^2 \operatorname{tr}\left(\Phi(\Phi^*\Phi)^{-1}\frac{\partial x^\star(y)}{\partial y}\right) .\end{aligned}$$

Note that if Φ is a Parseval tight frame operator, i.e. $\Phi\Phi^* = \text{Id}$, the prediction risk matches with the projection risk as well as the proposed GSURE estimates

$$\|\Pi x_0 - \Pi x^\star(y)\|^2 = \|\mu_0 - \mu^\star(y)\|^2 \quad \text{and} \quad \text{GSURE}_{\Pi}(y) = \text{GSURE}_{\Phi}(y) \ .$$

2.3. Analysis Sparsity Risk Estimation

Definition 4. *Let $\lambda \in \mathbb{R}_+^*$. We define the* λ-restricted transition space *as*

$$\mathcal{H}_{\cdot,\lambda} = \left\{ y \in \mathbb{R}^Q \setminus (y, \lambda) \in \mathcal{H} \right\}.$$

We first notice that even if $\mathcal{P}_\lambda(y)$ admits several solutions, all of them share the same image under Φ, see Section 5.3 for proof of this point. Hence, we denote without ambiguity $\Phi x^\star(y)$, where $x^\star(y)$ is a solution of $\mathcal{P}_\lambda(y)$. The following theorem is our third contribution.

Theorem 3. *Let $\lambda \in \mathbb{R}_+^*$. The λ-restricted transition space has a Lebesgue measure zero. Moreover, the mapping $y \mapsto \mu^\star(y)$ is of class C^∞ on $\mathbb{R}^Q \setminus \mathcal{H}_{\cdot,\lambda}$. For $y \notin \mathcal{H}_{\cdot,\lambda}$, there exists $x^\star$ a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds with J the D-cosupport of $x^\star$, and*

$$\frac{\partial \mu^\star(y)}{\partial y} = \Phi \Gamma^{[J]} \Phi^*. \tag{7}$$

For $y \notin \mathcal{H}_{\cdot,\lambda}$, we define $d(y) = \dim(\mathcal{G}_J)$ where J is the D-cosupport of any solution $x^\star$ such that (H_J) holds. We then obtain the following corollary as a consequence of Theorems 2 and 3.

Corollary 1. *With the notations of Section 2.2,*

$$\begin{aligned}
\mathrm{GSURE}_{\Phi}(y) =& \|y - \mu^{\star}(y)\|^2 - Q\sigma^2 + 2\sigma^2 d(y), \\
\mathrm{GSURE}_{\Pi}(y) =& \|x_{ML}(y) - \Pi x^{\star}(y)\|^2 - \sigma^2 \operatorname{tr}((\Phi\Phi^*)^{+}) + 2\sigma^2 \operatorname{tr}(\Pi\Gamma^{[J]}), \\
\mathrm{GSURE}_{\mathrm{Id}}(y) =& \|x_{ML}(y) - x^{\star}(y)\|^2 - \sigma^2 \operatorname{tr}((\Phi^*\Phi)^{-1}) + 2\sigma^2 \operatorname{tr}(\Gamma^{[J]}) \ .
\end{aligned}$$

Moreover, $d(y)$ is an unbiased estimator of the degrees of freedom of $\mathcal{P}_{\lambda}(y)$, i.e.

$$df(x^{\star}) = \mathbb{E}_w\left(d(y)\right).$$

2.4. Numerical Computation of the GSURE for sparse analysis estimators

The following proposition gives a way to compute efficiently the divergence term of sparse analysis estimators which boils down to solving a linear system.

Proposition 1. *One has*

$$\mathrm{dev}_A(y) = \mathbb{E}_Z(\langle \nu(Z),\, \Phi^* A^* A Z \rangle) \tag{8}$$

where $Z \sim \mathcal{N}(0, \mathrm{Id}_P)$, and where for any $z \in \mathbb{R}^P$, $\nu = \nu(z)$ solves the following linear system

$$\begin{pmatrix} \Phi^*\Phi & D_J \\ D_J^* & 0 \end{pmatrix} \begin{pmatrix} \nu \\ \tilde{\nu} \end{pmatrix} = \begin{pmatrix} \Phi^* z \\ 0 \end{pmatrix} \ . \tag{9}$$

In practice, the empirical mean estimator is replaced for the expectation in (8), hence giving

$$\frac{1}{k}\sum_{i=1}^{k} \langle \nu(z_i),\, \Phi^* A^* A z_i \rangle \overset{\mathrm{WLLN}}{\longrightarrow} \operatorname{tr}\left(A\Phi\Gamma^{[J]}\Phi^* A^*\right), \tag{10}$$

for k realizations z_i of Z. The numerical computation of $\nu(z_i)$ is achieved by solving the symmetric linear system (9) with a conjugate gradient solver.

3. Related Works

3.1. Local variations

The variations of the solution $x_{\lambda}(y)$ as a function of λ (Theorem 1, that also considers variations with respect to y) is already known in the synthesis case, see for instance [35, 15]. Our result also generalizes the work of [17] which studies the case of Φ overdetermined and develops an homotopy algorithm.

Theorem 3 is known to hold in the special case of synthesis regularization ($D = \mathrm{Id}$). It is proved in the overdetermined case in [36] and is extended to the general case in [37].

While this paper was ready for submission, it came to our attention that Tibshirani and Taylor [38, Theorem 3] recently and independently proved exactly the same result as our Theorem 3. Their proof uses a different approach, and in particular, they do not study directly the variations of $x_{\lambda}(y)$ as a function of y or λ (Theorem 1).

3.2. Generalized Stein Unbiased Risk Estimator

In the Gaussian context, our definitions of GSURE_{Π} and $\mathrm{GSURE}_{\mathrm{Id}}$ are equivalent, up to a constant which does not depends on λ, to the ones introduced in [34]. We have furthermore shown that both arises from a more general result given in Theorem 2. While the author of [34] imposes $x^*(y)$ to be a weakly differentiable function of $\Phi^* y/\sigma^2$, our definition does not rely on such an hypothesis, and it just requires that $y \mapsto \mu^\star(y)$ is weakly differentiable.

Indeed, let $u = \Phi^* y/\sigma^2$, and assume $y \mapsto x^*(y)$ is a weakly differentiable function of u, let say, $x^*(y) = z^*(u)$.

• When Φ is rank deficient, Eldar [34] defines an estimator of the projection risk given by

$$\mathrm{GSURE}_{\Pi}^{(\mathrm{Eldar})}(u) = \|\Pi x_0\|^2 + \|\Pi z^*(u)\|^2 - 2\,\langle z^*(u), x_{ML}(y)\rangle + 2\,\mathrm{tr}\left(\Pi \frac{\partial z^*(u)}{\partial u}\right).$$

Note that, since $u \mapsto z^*(u)$ is assumed to be weakly differentiable (and *a fortiori* defined without ambiguity), we have

$$\frac{\partial \Phi z^*(u)}{\partial u} = \Phi \frac{\partial z^*(u)}{\partial u}\ .$$

With the change of variable y to u, the following relation holds true

$$\sigma^2\,\mathrm{tr}\left((\Phi\Phi^*)^+ \frac{\partial \mu^*(y)}{\partial y}\right) = \sigma^2\,\mathrm{tr}\left((\Phi\Phi^*)^+ \frac{\partial \Phi z^*(u)}{\partial u}\frac{\partial u}{\partial y}\right) = \mathrm{tr}\left(\Pi \frac{\partial z^*(u)}{\partial u}\right)$$

and hence

$$\mathrm{GSURE}_{\Pi}(y) - \mathrm{GSURE}_{\Pi}^{(\mathrm{Eldar})}(y) = \|x_{ML}(y)\|^2 - \|\Pi x_0\|^2 - \sigma^2\,\mathrm{tr}\big((\Phi\Phi^*)^+\big)\ .$$

• When Φ has full rank, Eldar [34] defines an estimator of the estimation risk given by

$$\mathrm{GSURE}_{\mathrm{Id}}^{(\mathrm{Eldar})}(u) = \|x_0\|^2 + \|z^*(u)\|^2 - 2\,\langle z^*(u), x_{ML}(y)\rangle + 2\,\mathrm{tr}\left(\frac{\partial z^*(u)}{\partial u}\right)$$

With the change of variable y to u, we have

$$\sigma^2\,\mathrm{tr}\left(\Phi(\Phi^*\Phi)^{-1}\frac{\partial x^*(y)}{\partial y}\right) = \sigma^2\,\mathrm{tr}\left(\Phi(\Phi^*\Phi)^{-1}\frac{\partial z^*(u)}{\partial u}\frac{\partial u}{\partial y}\right) = \mathrm{tr}\left(\frac{\partial z^*(u)}{\partial u}\right)$$

and hence

$$\mathrm{GSURE}_{\mathrm{Id}}(y) - \mathrm{GSURE}_{\mathrm{Id}}^{(\mathrm{Eldar})}(y) = \|x_{\mathrm{ML}}(y)\|^2 - \|x_0\|^2 - \sigma^2\,\mathrm{tr}\big((\Phi^*\Phi)^{-1}\big)\ .$$

In both situations, the two estimators are asymptotically the same (their difference has an expectation of zero), and in particular, they are both unbiased. We can show that they have in general a different variance. However, since their difference does not depend on $x^*(.)$ and in particular on λ, both estimators lead to the same result when the purpose is to choose the optimal parameter λ.

In the context of deconvolution, GSURE_{Π} boils down to the unbiased estimator of the projection risk obtained in [39].

3.3. Numerical computation of the GSURE

In least square regression regularized by a smoothed penalization term, the DOF can be computed in closed-form [40]. For non-smooth sparse regularization, when closed-form expressions was not available, first attempts developed asymptotically unbiased estimators of the DOF [24]. Ye [41] and Shen and Ye [42] proposed a data perturbation technique to approximate the SURE when its closed-form expression is not available or numerically too expensive to compute. For denoising, a similar Monte Carlo approach has been used by Ramani *et al.* in [43] and applied to total-variation denoising, wavelet soft-thresholding, and Wiener filtering/smoothing splines.

Alternatively, an estimate can be obtained by formally differentiating an algorithm that computes or converges to the solution. Initially, it has been proposed by [27], and then refined in [44], to compute the GSURE of sparse synthesis regularization by differentiating the sequence of iterates of the forward-backward algorithm. Concurrently, a similar GSURE version has been proposed to non-iterative wavelet-vaguelet deconvolution [39]. We have recently proposed an extension of this methodology for proximal splitting algorithms solving a sparse analysis regularization that we applied to isotropic total-variation and $\ell^1 - \ell^2$ block sparsity [45].

The introduction of a closed-form expression of an unbiased estimate of the DOF for synthesis ℓ^1 regularization has open a new way to obtain unbiased estimates of the prediction risk [36, 37]. Our numerical computation of GSURE^A provides an estimate of various risk definitions of the solutions of analysis ℓ^1 regularization.

4. Examples

In this section, we exemplify the usefulness of our GSURE estimator which can serve as a basis for automatically tuning the value of λ. This is achieved by computing, from a single realization of the noise w, the parameter that minimizes the value of $\mathrm{GSURE}(y)$ for $y = \Phi x_0 + w$.

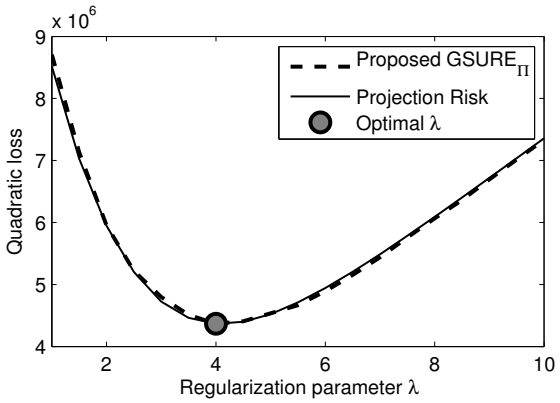
4.1. Computing Minimizers

Denoising. Although it is convex, solving $\mathcal{P}_\lambda(y)$ is rather challenging given its non-smoothness. In the case where $\Phi = \mathrm{Id}$, the functional of $\mathcal{P}_\lambda(y)$ is strictly convex, and one can compute its unique solution $x^\star$ by solving an equivalent dual problem [46]

$$x^\star = y + D\alpha^\star \quad \text{where} \quad \alpha^\star \in \operatorname*{argmin}_{\|\alpha\|_\infty \leqslant \lambda} \|y + D\alpha\|_2^2.$$

General Case. The proximity operator of $x \mapsto \|D^* x\|_1$ is not computable in closed-form for an arbitrary dictionary D. This precludes the use of popular iterative soft-thresholding (actually the forward-backward proximal splitting) without sub-iterating. We therefore appeal to more elaborate primal-dual splitting algorithm. We use in the numerical example the relaxed Arrow-Hurwicz

(a) y (b) x^* at the optimal λ (c) x_0



(d)

Figure 1: Illustration of the optimal selection of λ in a super-resolution problem ($Q/N = 0.5$) with anisotropic total variation regularization ($D^* = \nabla$). (a) The observed image y. (b) The solution x^* at the optimal λ. (c) The underlying true image x_0. (d) Projection risk and its GSURE estimate obtained using $k = 1$ random realization.

algorithm as revitalized in [47]. This algorithm achieves full splitting where all operators are applied separately: proximity operators of $\frac{1}{2}\|\cdot\|^2$ and $\lambda\|\cdot\|_1$, Φ, D and their adjoints.

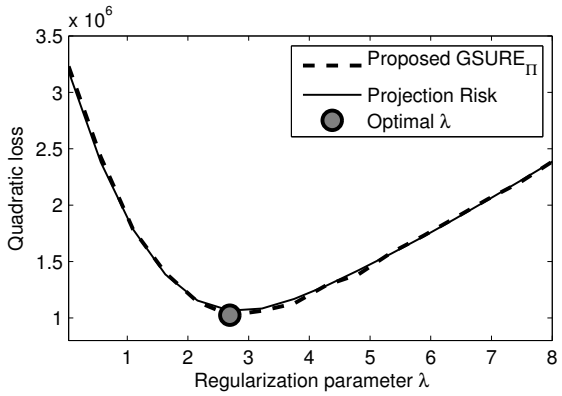
$$\min_{x\in\mathbb{R}^N} F(K(x)) \quad \text{where} \quad \left\{ \begin{array}{l} F(g,u) = \frac{1}{2}\|y-g\|_2^2 + \lambda\|u\|_1 \\ K(x) = (\Phi x, D^* x). \end{array} \right.$$

Several others [48, 49] algorithms exist.

4.2. Computing GSURE Estimates

Total Variation Regularization. In this example, Φ is a vertical sub-sampling operator suppressing one line over two (hence $Q/N = 0.5$). The noise level has been set such that the input image y has a PSNR of 27.78 dB. The regularization is an anisotropic total variation regularization ($D^* = \nabla$) where ∇ is the gradient operator. Fig. 1.d depicts the projected risk and its GSURE_Π estimate with $k = 1$ as a function of λ. The curves are indeed unimodal and coincide even with $k = 1$ and a single noise realization. Consequently, GSURE_Π provides a

(a) x_{ML} (b) x^* at the optimal λ (c) x_0



(d)

Figure 2: Illustration of optimal selection of λ in a compressed sensing problem ($Q/N = 0.5$) by reducing the correlations with the detailed atoms of a multi-scale shift-invariant Haar dictionnary. (a) The least square estimate x_{ML}. (b) The solution x^* at the optimal λ. (c) The underlying true image x_0. (d) Projection risk and its GSURE estimate obtained using $k = 1$ random realization.

high-quality estimate of λ minimizing the risk. A close in on the central area of the degraded, over-sampled (using the optimal λ), and true images is shown in Fig. 1(a)-(c) for visual inspection of the restoration quality.

Sparse Analysis Regularization. We consider in this example a compressed sensing setting where Φ is a random partial DCT measurement matrix with an under-sampling ratio $Q/N = 0.5$. The input image y has a PSNR set to 27.50 dB. The regularization is $D = \Psi$ where Ψ^* is a $6N \times N$ matrix whose columns compute the detailed coefficients of a multi-scale shift-invariant Haar decomposition (with 3 scales in horizontal and vertical directions). We estimate GSURE_Π with $k = 1$. The results observed on the super-resolution example are confirmed in this compressed sensing experiment both visually and qualitatively, see Fig. 2.

5. Proofs

This section details the proofs of Theorems 1–3. The objective function $\mathcal{L}_{y,\lambda}$ minimized in $\mathcal{P}_\lambda(y)$ is

$$\mathcal{L}_{y,\lambda}(x) = \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \|D^* x\|_1.$$

We recall that we suppose that condition (H_0) holds in every statements. The following lemma, which is at the heart of the proofs of our contributions, details the first order optimality conditions for the analysis variational problem $\mathcal{P}_\lambda(y)$.

Lemma 1. *A vector $x^\star$ is a solution of $\mathcal{P}_\lambda(y)$ if, and only if, there exists $\sigma \in \mathbb{R}^{|J|}$, where J is the D-cosupport of $x^\star$, such that*

$$\sigma \in \Sigma_{y,\lambda}(x^\star) \tag{11}$$

where $I = J^c$ the D-support,

$$\Sigma_{y,\lambda}(x^\star) = \left\{ \sigma \in \mathbb{R}^{|J|} \setminus \Phi^*(\Phi x^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \textit{and} \quad \|\sigma\|_\infty \leqslant 1 \right\}. \tag{12}$$

and $s = \operatorname{sign}(D^* x^\star)$.

Proof. The subdifferential ∂F of a real valued convex lower semicontinuous function $F : \mathbb{R}^N \to \mathbb{R}$ is the multifunction defined by

$$\partial F(x_0) = \left\{ g \in \mathbb{R}^N \setminus \forall x \in \mathbb{R}^N, f(x) \geqslant f(x_0) + \langle g,\, x - x_0 \rangle \right\}.$$

Note that x_0 is a minimum of F if, and only if, $0 \in \partial F(x_0)$. Indeed, if $0 \in \partial F(x_0)$, then for every $x \in \mathbb{R}^N, F(x) \geqslant F(x_0)$, meaning that x_0 is a minimum of F over $\mathbb{R}^N$. The subdifferential of $\mathcal{L}_{y,\lambda}(x)$ is

$$\partial \mathcal{L}_{y,\lambda}(x) = \left\{ \Phi^*(\Phi x - y) + \lambda D u \setminus u \in \mathbb{R}^N : u_I = \operatorname{sign}(D^* x)_I \text{ and } \|u_J\|_\infty \leqslant 1 \right\}.$$

Hence $0 \in \partial \mathcal{L}_{y,\lambda}(x)$ is equivalent to the existence of $u \in \mathbb{R}^N$ such that $u_I = \operatorname{sign}(D^* x)_I$ and $\|u_J\|_\infty \leqslant 1$ satisfiyng

$$\Phi^*(\Phi x - y) + \lambda D u = 0.$$

Defining $\sigma = u_J$, it is equivalent to the existence of $\sigma \in \Sigma_{y,\lambda}(x)$ with $\|\sigma\|_\infty \leqslant 1$. □

5.1. Proof of Theorem 1

The proof of Theorem 1 is done in three steps. First, we prove Lemma 2 which gives an implicit equation satisfied by a solution of $\mathcal{P}_\lambda(y)$. Then, we prove Lemma 3. Finally, we proves Theorem 1.

The following lemma gives an implicit equation satisfied by a solution $x^\star$ of the problem $\mathcal{P}_\lambda(y)$. Note that $\mathcal{P}_\lambda(y)$ may have other solutions.

Lemma 2. *Let $x^\star$ a solution of $\mathcal{P}_\lambda(y)$. Let I be the D-support and J the D-cosupport of $x^\star$ and $s = \operatorname{sign}(D^* x^\star)$. We suppose that (H_J) holds. Then, $x^\star$ satisfies*

$$x^\star = \Gamma^{[J]} \Phi^* y - \lambda \Gamma^{[J]} D_I s_I. \tag{13}$$

Proof. Using the first order condition (Lemma 1) there exists $\sigma \in \Sigma_{y,\lambda}(x^\star)$ satisfying

$$\Phi^*(\Phi x^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0. \tag{14}$$

By definition, one has $x^\star \in \mathcal{G}_J$ so $x^\star \in (\operatorname{Im} D_J)^\perp$. Hence, we can write $x^\star = U\alpha$. Since $U^* D_J = 0$, multiplying equation (14) on the left by U^*, we get

$$U^* \Phi^*(\Phi U \alpha - y) + \lambda U^* D_I s_I = 0.$$

Since $U^* \Phi^* \Phi U$ is invertible, we conclude. □

Lemma 3. *Let $y \in \mathbb{R}^P$ and let J a D-cosupport such that (H_J) holds, and $I = J^c$. Suppose $\hat{x}^\star$ satisfies*

$$\hat{x}^\star = \Gamma^{[J]} \Phi^* y - \lambda \Gamma^{[J]} D_I s_I.$$

where $s = \operatorname{sign}(D^* \hat{x}^\star)$. Then, $\hat{x}^\star$ is a solution of $\mathcal{P}_\lambda(y)$ if, and only if, there exists σ satisfying one of the following conditions

$$\sigma - \Omega^{[J]} s_I + \frac{1}{\lambda} \Pi^{[J]} y \in \operatorname{Ker} D_J \quad \textit{and} \quad \|\sigma\|_\infty \leqslant 1, \tag{15}$$

or equivalently,

$$\tilde{\Pi}^{[J]} y - \lambda \tilde{\Omega}^{[J]} s_I + \lambda D_J \sigma = 0 \quad \textit{and} \quad \|\sigma\|_\infty \leqslant 1, \tag{16}$$

where $\tilde{\Omega}^{[J]} = (\Phi^* \Phi \Gamma^{[J]} - \mathrm{Id}) D_I$, $\tilde{\Pi}^{[J]} = \Phi^*(\Phi \Gamma^{[J]} \Phi^* - \mathrm{Id})$, $\Omega^{[J]} = D_J^+ \tilde{\Omega}^{[J]}$ and $\Pi^{[J]} = D_J^+ \tilde{\Pi}^{[J]}$.

Proof. Remark that $\hat{x}^\star$ is an element of $\mathcal{G}_J$. According to Lemma 1, $\hat{x}^\star$ is a solution of $\mathcal{P}_\lambda(y)$ if, and only if, there exists $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$ such that

$$\Phi^*(\Phi \hat{x}^\star - y) + \lambda D_I s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1.$$

Since (H_J) holds, one can define $\Gamma^{[J]}$. We use the implicit equation (13),

$$\Phi^*(\Phi \Gamma^{[J]} \Phi^* y - \lambda \Phi \Gamma^{[J]} D_I s_I - y) + \lambda D_I s_I + \lambda D_J \sigma = 0.$$

Factorizing the term in front of y and s_I, one has

$$\Phi^*(\Phi \Gamma^{[J]} \Phi^* - \mathrm{Id}) y - \lambda (\Phi^* \Phi \Gamma^{[J]} - \mathrm{Id}) D_I s_I + \lambda D_J \sigma = 0.$$

which proves that

$$\tilde{\Pi}^{[J]} y - \lambda \tilde{\Omega}^{[J]} s_I + \lambda D_J \sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1,$$

One has $U^*\tilde{\Omega} = 0$ and thus one remarks that $\Omega^{[J]} = D_J^+\tilde{\Omega}^{[J]}$. Similarly, we define $\tilde{\Pi}^{[J]}$ such that $\Pi^{[J]} = D_J^+\tilde{\Pi}^{[J]}$. Hence, the existence of $\sigma \in \Sigma_{y,\lambda}(\hat{x}^\star)$ such that $\|\sigma\|_\infty \leqslant 1$ is equivalent to

$$D_J\sigma = D_J\Omega^{[J]}s_I - \frac{1}{\lambda}D_J\Pi^{[J]}y \quad \text{where} \quad \|\sigma\|_\infty \leqslant 1,$$

which in turn is equivalent to

$$\sigma - \Omega^{[J]}s_I + \frac{1}{\lambda}\Pi^{[J]}y \in \operatorname{Ker} D_J \quad \text{where} \quad \|\sigma\|_\infty \leqslant 1.$$

□

We now prove Theorem 1.

Proof of Theorem 1. Let $(y,\lambda) \notin \mathcal{H}$. By construction of the vector $\hat{x}_{\bar{\lambda}}(\bar{y})$ one has $D_J^*\hat{x}_{\bar{\lambda}}(\bar{y}) = 0$. So for $(\bar{y},\bar{\lambda})$ close enough from (y,λ), one has

$$\operatorname{sign}(D^*\hat{x}_{\bar{\lambda}}(\bar{y})) = \operatorname{sign}(D^*x^\star(y)).$$

Since $x^\star$ is a solution of $\mathcal{P}_\lambda(y)$, using Lemmas 2 and 3, there exists σ such that

$$\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_J\sigma = 0 \quad \text{and} \quad \|\sigma\|_\infty \leqslant 1. \tag{17}$$

We split $J = K \cup L$, $K \cap L = \emptyset$ such that $\|\sigma_K\|_\infty = 1$ and $\|\sigma_L\|_\infty < 1$. We first suppose that $\operatorname{Im}\tilde{\Pi}^{[J]} \subseteq \operatorname{Im} D_L$. To prove that $\hat{x}_{\bar{\lambda}}(\bar{y})$ is solution to $\mathcal{P}_{\bar{\lambda}}(\bar{y})$ we show that there exists $\bar{\sigma}$ such that $\|\bar{\sigma}\|_\infty \leqslant 1$ and

$$\tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_K\bar{\sigma}_K + \bar{\lambda}D_L\bar{\sigma}_L = 0.$$

We impose that $\bar{\sigma}_K = \sigma_K$ and we introduce $\bar{\sigma}_L$ as

$$\bar{\sigma}_L = \sigma_L - \frac{1}{\lambda}D_L^+\tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right).$$

Hence,

$$\begin{aligned}
& \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_J\bar{\sigma} \\
= \;& \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_K\sigma_K + \bar{\lambda}D_L\sigma_L \\
& -D_LD_L^+\frac{\bar{\lambda}}{\lambda}\tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right) \\
= \;& \underbrace{\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_K\sigma_K + \lambda D_L\sigma_L}_{=0} \\
& -\tilde{\Pi}^{[J]}(y-\bar{y}) + (\lambda-\bar{\lambda})\tilde{\Omega}^{[J]}s_I - (\lambda-\bar{\lambda})D_K\sigma_K - (\lambda-\bar{\lambda})D_L\sigma_L \\
& -D_LD_L^+\frac{\bar{\lambda}}{\lambda}\tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right)
\end{aligned}$$

Since $\operatorname{Im}\tilde{\Pi}^{[J]} \subseteq \operatorname{Im} D_L$, there exists u such that

$$\tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right) = D_L u.$$

By property of Moore-Penrose pseudo-inverse,

$$D_L D_L^+ D_L u = D_L u = \tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right).$$

Hence,

$$\begin{aligned} & \tilde{\Pi}^{[J]}\bar{y} - \bar{\lambda}\tilde{\Omega}^{[J]}s_I + \bar{\lambda}D_J\bar{\sigma} \\ = \; & \frac{\bar{\lambda}-\lambda}{\lambda}\left[\tilde{\Pi}^{[J]}y - \lambda\tilde{\Omega}^{[J]}s_I + \lambda D_K\sigma_K + \lambda D_L\sigma_L\right] \\ = \; & 0. \end{aligned}$$

If $(\bar{y}, \bar{\lambda})$ is close enough from (y, λ), one has

$$\|\bar{\sigma}_L\|_\infty = \|\sigma_L - \frac{1}{\lambda}D_L^+\tilde{\Pi}^{[J]}\left(\frac{\lambda-\bar{\lambda}}{\bar{\lambda}}y + \frac{\lambda}{\bar{\lambda}}(\bar{y}-y)\right)\|_\infty \leqslant 1,$$

i.e $\hat{x}_{\bar{\lambda}}(\bar{y})$ is solution of $\mathcal{P}_{\bar{\lambda}}(\bar{y})$. Suppose now that $\operatorname{Im}\tilde{\Pi} \not\subseteq \operatorname{Im} D_L$. Then remark that projecting (17) on $\mathcal{G}_L$ shows that

$$P_{\mathcal{G}_L}\tilde{\Pi}^{[J]}y = \lambda(\tilde{\Omega}^{[J]}s_{J^c} + D_K s_K),$$

which is a contradiction of $(y, \lambda) \notin \mathcal{H}$. □

5.2. Proof of Theorem 2

First we recall Stein's lemma and then we prove Theorem 2.

Lemma 4 (Stein's lemma)**.** *Let $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2\mathrm{Id}_Q)$. Assume $g : y \mapsto g(y)$ is weakly differentiable, then*

$$\mathbb{E}_w \langle w, g(y)\rangle = 2\sigma^2\mathbb{E}_w \operatorname{tr}\left(\frac{\partial g(y)}{\partial y}\right)$$

A proof of the lemma can be found in [22].

Proof of Theorem 2. Let $A \in \mathbb{R}^{M\times Q}$ and $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2\mathrm{Id}_Q)$ Using the decomposition $Ay = A\Phi x_0 + Aw$, one has

$$\begin{aligned} \mathbb{E}_w\|Ay - A\Phi x^\star(y)\|^2 &= \mathbb{E}_w\|A\Phi x_0 + Aw\|^2 - 2\mathbb{E}_w\langle A\Phi x_0 + Aw, A\Phi x^\star(y)\rangle \\ &\quad + \mathbb{E}_w\|A\Phi x^\star(y)\|^2 \\ &= \mathbb{E}_w\|A\Phi x_0\|^2 + \sigma^2\operatorname{tr}(A^*A) - 2\mathbb{E}_w\langle A\Phi x_0, A\Phi x^\star(y)\rangle \\ &\quad - 2\mathbb{E}_w\langle w, A^*A\Phi x^\star(y)\rangle + \mathbb{E}_w\|A\Phi x^\star(y)\|^2 \ . \end{aligned}$$

Assuming $y \mapsto \Phi x^\star(y)$ is weakly differentiable, we have

$$\frac{\partial A^* A \Phi x^\star(y)}{\partial y} = A^* A \frac{\partial \Phi x^\star(y)}{\partial y} \ .$$

and Lemma 4 gives

$$\mathbb{E}_w \|Ay - A\Phi x^\star(y)\|^2 = \mathbb{E}_w \|A\Phi x_0 - A\Phi x^\star(y)\|^2 + \sigma^2 \operatorname{tr}(A^* A) - 2\sigma^2 \mathbb{E}_w \operatorname{tr}\left(A^* A \frac{\partial \Phi x^\star(y)}{\partial y} \right) \ .$$

□

5.3. Proof of Theorem 3

The proof is done in four steps. First, we prove that $\mu(y)$ is well-defined. Then, we prove that there exists a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds. Finally, we prove that $\operatorname{div}(\mu)(y) = \dim \mathcal{G}_J$.

We first proves that even if $\mathcal{P}_\lambda(y)$ admits several solutions, all of them share the same image under Φ.

Lemma 5. *If x_1 and x_2 are two solutions of $\mathcal{P}_\lambda(y)$, then $\Phi x_1 = \Phi x_2$.*

Proof. Let x_1, x_2 be two solutions of $\mathcal{P}_\lambda(y)$ and $\Phi x_1 \neq \Phi x_2$. We define $x_3 = \frac{1}{2}(x_1 + x_2)$. Since the function $u \mapsto \|y - u\|^2$ is strictly convex, one has the following inequality

$$\frac{1}{2}\|y - \Phi x_3\|^2 < \frac{1}{2}\left(\frac{1}{2}\|y - \Phi x_1\|^2 + \frac{1}{2}\|y - \Phi x_2\|^2 \right).$$

Applying triangle inequality for the ℓ^1 norm gives

$$\|D^* x_3\|_1 \leqslant \|D^* x_1\|_1 + \|D^* x_2\|_1.$$

Hence, $\mathcal{L}_{y,\lambda}(x_3) < \mathcal{L}_{y,\lambda}(x_1)$ which is a contradiction with x_1 being a solution of the problem $\mathcal{P}_\lambda(y)$. □

Lemma 6. *There exists $x^\star$ a solution of $\mathcal{P}_\lambda(y)$ such that (H_J) holds, where J is the D-cosupport of $x^\star$.*

Proof. Let $x^\star$ be a solution of $\mathcal{P}_\lambda(y)$. Suppose (H_J) does not hold. Our strategy is to prove that there exists a solution of D-support strictly included in $I = J^c$.

Since (H_J) does not hold, there exists $z \in \operatorname{Ker} \Phi$ with $z \neq 0$ and $D_J^* z = 0$. We define for every $t \in \mathbb{R}$, the vector $v_t = x^\star + tz$. Denote $\mathcal{B}$ the subset of $\mathbb{R}$ defined by

$$\mathcal{B} = \{ t \in \mathbb{R} \setminus \operatorname{sign}(D^* v_t) = \operatorname{sign}(D^* x^\star) \} ,$$

The set $\mathcal{B}$ is a non empty set, $0 \in \mathcal{B}$ and convex from its definition. Moreover for all $t \in \mathcal{B}$, $\partial \mathcal{L}_{y,\lambda}(v_t) = \partial \mathcal{L}_{y,\lambda}(x^\star)$, it follows that for all $t \in \mathcal{B}$, v_t is a solution of $\mathcal{P}_\lambda(y)$. As a consequence,

$$\forall t \in \mathcal{B}, \quad \Phi v_t = \Phi x^\star \quad \text{and} \quad \|D^* v_t\|_1 = \|D^* x^\star\|_1.$$

Since $\lim_{|t|\to\infty} \|D^* v_t\|_1 = +\infty$, the set $\mathcal{B}$ is bounded. Hence, $\mathcal{B}$ is an open interval of $\mathbb{R}$ which contain 0, i.e there exist $t_1, t_0 \in \mathbb{R}$ such that

$$\mathcal{B} =]t_1, t_0[\quad \text{where} \quad -\infty < t_1 < 0 \quad \text{and} \quad 0 < t_0 < +\infty.$$

Since $t_0 \notin \mathcal{B}$, the D-support of v_{t_0} is strictly included in I. Moreover by continuity,

$$\Phi v_{t_0} = \Phi x^\star \quad \text{and} \quad \|D^* v_{t_0}\|_1 = \|D^* x^\star\|_1.$$

Hence, v_{t_0} is a solution of $\mathcal{P}_\lambda(y)$ of D-support strictly included in I.

Iterating this argument for $x^\star = v_{t_0}$ shows that there exists a solution such that (H_J) holds. □

We now prove the Theorem 3 starting with a lemma on the measure of $\mathcal{H}_{\cdot,\lambda}$.

Lemma 7. *Let $J \subset \{1, \cdots, P\}$ such that (H_J) holds, K a subset of J such that $\operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}$, $s_{J^c} \in \{-1,1\}^{|J^c|}$ and $s_K \in \{-1,1\}^{|K|}$. If $\operatorname{Im} \tilde{\Pi}^{[J]}$ is not included in $\operatorname{Im} D_{K^c}$ then $\mathcal{H}_{J,K,s_{J^c},\sigma_K}$ is an affine space of $\mathbb{R}^Q \times \mathbb{R}$ and different from $\mathbb{R}^Q \times \mathbb{R}$. Moreover, $\mathcal{H}$ has a Lebesgue measure zero and for every $\lambda \in \mathbb{R}_+^*$ $\mathcal{H}_{\cdot,\lambda}$ has a Lebesgue measure zero.*

Proof. Consider $J \subset \{1, \cdots, P\}$ such that (H_J) holds, K a subset of J such that $\operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}$, $s_{J^c} \in \{-1,1\}^{|J^c|}$ and $s_K \in \{-1,1\}^{|K|}$. The following set

$$\mathcal{H}_{J,K,s_{J^c},s_K} = \left\{ (y,\lambda) \in \mathbb{R}^Q \times \mathbb{R} \setminus P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} y = \lambda(\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \right\},$$

is a vector subspace of $\mathbb{R}^Q \times \mathbb{R}$. Indeed, let $(y_1, \lambda_1), (y_2, \lambda_2) \in \mathcal{H}_{J,K,s_{J^c},s_K}$ and $\mu \in \mathbb{R}$. Hence,

$$P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]}(y_1 + \mu y_2) = (\lambda_1 + \mu\lambda_2)(\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K).$$

Moreover $(0_Q, 0) \in \mathcal{H}_{J,K,s_{J^c},s_K}$. Each $\mathcal{H}_{J,K,s_{J^c},s_K}$ is different from $\mathbb{R}^Q \times \mathbb{R}$. Indeed, $(y,\lambda) \in \mathcal{H}_{J,K,s_{J^c},s_K}$ is equivalent to

$$\begin{pmatrix} P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} & 0 \\ 0 & -\lambda \mathrm{Id} \end{pmatrix} \begin{pmatrix} y \\ \tilde{\Omega}^{[J]} s_{J^c} + D_K s_K \end{pmatrix} = 0.$$

Particularly, we fixed λ. If every $y \in \mathbb{R}^Q$ is solution of this system, the matrix $P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]}$ is invertible, which is impossible since $P_{\mathcal{G}_{K^c}}$ is an orthogonal projection on a strict subspace of $\mathbb{R}^Q$. Since $\mathcal{H}$ is a finite unition of $\mathcal{H}_{J,K,s_{J^c},s_K}$ all different from $\mathbb{R}^Q \times \mathbb{R}$, $\mathcal{H}$ has a Lebesgue measure zero. Remark that $\mathcal{H}_{\cdot,\lambda}$ is included in

$$\tilde{\mathcal{H}}^\lambda = \bigcup_{\substack{J \subset \{1,\cdots,P\} \\ (H_J) \text{ holds}}} \bigcup_{\substack{K \subset J \\ \operatorname{Im} \tilde{\Pi}^{[J]} \not\subset \operatorname{Im} D_{K^c}}} \bigcup_{s_{J^c} \in \{-1,1\}^{|J^c|}} \bigcup_{s_K \in \{-1,1\}^{|K|}} \tilde{\mathcal{H}}_{J,K,s_{J^c},s_K},$$

where

$$\tilde{\mathcal{H}}^\lambda_{J,K,s_{J^c},s_K} = \left\{ y \in \mathbb{R}^Q \setminus P_{\mathcal{G}_{K^c}} \tilde{\Pi}^{[J]} y = \lambda(\tilde{\Omega}^{[J]} s_{J^c} + D_K s_K) \right\},$$

Similarly to $\mathcal{H}_{J,K,s_{J^c},s_K}$, we prove that each $\tilde{\mathcal{H}}^\lambda_{J,K,s_{J^c},s_K}$ is a strict affine subspace of $\mathbb{R}^Q$. Hence, $\tilde{\mathcal{H}}^\lambda$ has a Lebesgue measure zero, and so does $\mathcal{H}_{\cdot,\lambda}$. □

Next, we prove the theorem 3.

Proof of Theorem 3. Using Lemma 6, there exists a solution $x^\star(y)$ of $\mathcal{P}_\lambda(y)$ such that (H_J) holds. We consider this solution. Using Theorem 1 for $\bar{y}$ close enough from y one has

$$\Phi\hat{x}_{\bar{\lambda}}(\bar{y}) = \Phi\Gamma^{[J]}\Phi^*\bar{y} - \lambda\Phi\Gamma^{[J]}D_I s_I.$$

where J is the D-cosupport of $x^\star(y)$. Remark that $\Phi\hat{x}_{\bar{\lambda}}(\bar{y})$ can be written as $\Phi\Gamma^{[J]}\Phi^*\bar{y} + r$ and $r \in \mathbb{R}^P$ is a constant vector. Hence,

$$\frac{\partial\Phi\hat{x}_{\bar{\lambda}}(\bar{y})}{\partial y} = \Phi\Gamma^{[J]}\Phi^*.$$

Moreover, using Lemma 7, $\mathcal{H}_{\cdot,\lambda}$ has a Lebesgue measure zero. □

We now prove the corrolary.

Proof of Corollary 1. Let $\lambda \in \mathbb{R}_+^*$. Using Lemma 7, $\mathcal{H}_{\cdot,\lambda}$ has a Lebesgue measure zero. Hence, $y \mapsto \Phi x^\star(y)$ is differentiable almost everywhere and we can apply Theorem 2.

Remark that $\mathrm{dev}_\Phi(y) = \mathrm{tr}(\Phi\Gamma^{[J]}\Phi^*)$ and $V = \Phi\Gamma^{[J]}\Phi^*$ is the orthogonal projector on $\mathrm{Im}(V) = \ker(V)^\perp$, so that $\mathrm{tr}(V) = \dim(\mathrm{Im}(V))$. Since Φ is injective on $\mathcal{G}_J$, one has $\dim(\mathrm{Im}(V)) = \dim(\mathcal{G}_J)$.

Moreover, $\mathrm{cov}(y, \Phi x^\star(y)) = \mathbb{E}_w \langle w, \Phi x^\star(y)\rangle$. By definition of the degrees of freedom, using Lemma 4:

$$df(\Phi x^\star) = \sum_{i=1}^{Q} \frac{\mathrm{cov}(y, \Phi x^\star(y))}{\sigma^2} = \mathbb{E}_w\left(\mathrm{tr}(\Phi\Gamma^{[J]}\Phi^*)\right) .$$

□

Proof of Proposition 1. We have

$$\mathrm{tr}\left(A\Phi\Gamma^{[J]}\Phi^*A^*\right) = \mathrm{tr}\left(\Phi\Gamma^{[J]}\Phi^*A^*A\right).$$

Hence denoting $\nu(z) = \Gamma^{[J]}\Phi^*z$, and using the fact that for any matrix U, $\mathrm{tr}(U) = \mathbb{E}_Z(\langle Z, UZ\rangle)$, we arrive at (8).

We then use the fact that $\Gamma^{[J]}\Phi^*$, the inverse of Φ on $\mathcal{G}_J$, is the mapping that solves the following linearly constrained least-squares problem

$$\Gamma^{[J]}\Phi^*z = \underset{h\in\mathcal{G}_J}{\mathrm{argmin}} \|\Phi h - z\|^2.$$

The closed-form solution to this problem is given by (9). □

6. Conclusion

This paper studies the local behavior of solutions to sparse analysis regularized inverse problems of the form $\mathcal{P}_\lambda(y)$. We proved that the minimizers $x^\star$ of $\mathcal{P}_\lambda(y)$ are piecewise affine functions with respect to the observations y and the regularization parameter λ. This local affine parametrization is completely characterized by the D-support I of $x^\star$, i.e. the set of atoms in D with non-zero correlations with $x^\star$. Consequently, for y outside a set of zero Lebesgue measure, the first-order variations of $x^\star$ with respect to y is obtained in closed-form.

We capitalized on these results to objectively and automatically choose the regularization parameter λ. Toward this goal, a unified framework to unbiasedly estimate several risk measures is proposed through the GSURE. This encompasses several special cases such as unbiased estimation of the prediction, the projection and the estimation risk. An efficient algorithm is designed to compute this general estimator (GSURE) in the context of sparse analysis reconstruction. Illustrations on different imaging inverse problems exemplify the potential applicability of our theoretical findings.

References

[1] A. Kirsch, An introduction to the mathematical theory of inverse problems, volume 120, Springer Verlag, 2011.

[2] J. Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms: Fundamentals, volume 305, Springer-Verlag, 1996.

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

[4] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, T. Pock, Theoretical Foundations and Numerical Methods for Sparse Recovery, De Gruyter, pp. 263–340.

[5] M. Nikolova, Local strong homogeneity of a regularized estimator, SIAM Journal on Applied Mathematics 61 (2000) 633–658.

[6] R. Tibshirani, Regression shrinkage and selection via the Lasso, Journal of the Royal Statistical Society. Series B. Methodological 58 (1996) 267–288.

[7] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM journal on scientific computing 20 (1999) 33–61.

[8] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, Inverse Problems 23 (2007) 947–968.

[9] S. Nam, M. Davies, M. Elad, R. Gribonval, The cosparse analysis model and algorithms, 2011. Preprint arxiv-1106.4987v1.

[10] S. Vaiter, G. Peyré, C. Dossal, M. Fadili, Robust Sparse Analysis Regularization, Technical Report, Preprint Hal-00627452, 2011.

[11] J. Bonnans, A. Shapiro, Perturbation analysis of optimization problems, Springer Verlag, 2000.

[12] B. Mordukhovich, Sensitivity analysis in nonsmooth optimization, Theoretical Aspects of Industrial Design (D. A. Field and V. Komkov, eds.), SIAM Volumes in Applied Mathematics 58 (1992) 32–46.

[13] W. Schirotzek, L. MyiLibrary, Nonsmooth analysis, Springer Berlin, 2007.

[14] M. Osborne, B. Presnell, B. Turlach, On the lasso and its dual, Journal of Computational and Graphical statistics (2000) 319–337.

[15] M. Osborne, B. Presnell, B. Turlach, A new approach to variable selection in least squares problems, IMA journal of numerical analysis 20 (2000) 389.

[16] D. Malioutov, M. Cetin, A. Willsky, Homotopy continuation for sparse signal representation, in: Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on, volume 5, IEEE, pp. 733–736.

[17] R. Tibshirani, J. Taylor, The solution path of the generalized Lasso, The Annals of Statistics 39 (2011) 1335–1371.

[18] B. Efron, How biased is the apparent error rate of a prediction rule?, Journal of the American Statistical Association 81 (1986) 461–470.

[19] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Second international symposium on information theory, volume 1, Springer Verlag, pp. 267–281.

[20] G. Schwarz, Estimating the dimension of a model, The annals of statistics 6 (1978) 461–464.

[21] G. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics (1979) 215–223.

[22] C. Stein, Estimation of the mean of a multivariate normal distribution, The Annals of Statistics 9 (1981) 1135–1151.

[23] K.-C. Li, From Stein's unbiased risk estimates to the method of generalized cross validation, Ann. Statist. 13 (1985) 1352–1377.

[24] B. Efron, The estimation of prediction error, Journal of the American Statistical Association 99 (2004) 619–632.

[25] D. Donoho, I. Johnstone, Adapting to Unknown Smoothness Via Wavelet Shrinkage., Journal of the American Statistical Association 90 (1995) 1200–1224.

[26] T. Blu, F. Luisier, The SURE-LET approach to image denoising, IEEE Trans. Image Process. 16 (2007) 2778–2786.

[27] C. Vonesch, S. Ramani, M. Unser, Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, in: ICIP, IEEE, pp. 665–668.

[28] T. Cai, H. Zhou, A data-driven block thresholding approach to wavelet estimation, The Annals of Statistics 37 (2009) 569–595.

[29] D. Van De Ville, M. Kocher, SURE-based Non-Local Means, IEEE Signal Process. Lett. 16 (2009) 973–976.

[30] V. Duval, J.-F. Aujol, Y. Gousseau, A bias-variance approach for the non-local means, SIAM Journal Imaging Sci. 4 (2011) 760–788.

[31] C.-A. Deledalle, V. Duval, J. Salmon, Non-local Methods with Shape-Adaptive Patches (NLM-SAP), Journal of Mathematical Imaging and Vision (2011) 1–18.

[32] J. Rice, Choice of smoothing parameter in deconvolution problems, Contemporary Mathematics 59 (1986) 137–151.

[33] L. Desbat, D. Girard, The 'minimum reconstruction error' choice of regularization parameters: Some more efficient methods and their application to deconvolution problems, SIAM J. Sci. Comput 16 (1995) 1387–1403.

[34] Y. C. Eldar, Generalized SURE for exponential families: Applications to regularization, IEEE Transactions on Signal Processing 57 (2009) 471–481.

[35] D. Donoho, Y. Tsaig, Fast solution of ℓ^1-norm minimization problems when the solution may be sparse, IEEE Transactions on Information Theory 54 (2008) 4789–4812.

[36] H. Zou, T. Hastie, R. Tibshirani, On the "degrees of freedom" of the lasso, The Annals of Statistics 35 (2007) 2173–2192.

[37] M. Kachour, C. Dossal, M. Fadili, G. Peyré, C. Chesneau, The degrees of freedom of penalized l1 minimization, Technical Report, Preprint Hal-00638417, 2011. Submitted.

[38] R. J. Tibshirani, J. Taylor, Degrees of Freedom in Lasso Problems, Technical Report, arXiv:1111.0653, 2012.

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

[40] V. Solo, A SURE-fired way to choose smoothing parameters in ill-conditioned inverse problems, in: IEEE Int. Conf. Image Process. (ICIP), volume 3, IEEE, pp. 89–92.

[41] J. Ye, On measuring and correcting the effects of data mining and model selection, Journal of the American Statistical Association (1998) 120–131.

[42] X. Shen, J. Ye, Adaptive model selection, Journal of the American Statistical Association 97 (2002) 210–221.

[43] S. Ramani, T. Blu, M. Unser, Monte-Carlo SURE: a black-box optimization of regularization parameters for general denoising algorithms, IEEE Trans. Image Process. 17 (2008) 1540–1554.

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

[45] C. Deledalle, S. Vaiter, G. Peyré, J. Fadili, C. Dossal, Proximal Splitting Derivatives for Risk Estimation, Technical Report, Preprint Hal-00670213, 2012.

[46] A. Chambolle, An algorithm for total variation minimization and applications, Journal of Mathematical Imaging and Vision 20 (2004) 89–97.

[47] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (2011) 120–145.

[48] H. Raguet, J. Fadili, G. Peyré, Generalized Forward-Backward Splitting, Technical Report, Preprint Hal-00613637, 2011.

[49] N. Pustelnik, C. Chaux, J.-C. Pesquet, Parallel proXimal algorithm for image restoration using hybrid regularization, IEEE Transactions on Image Processing 20 (2011) 2450–2462.