Appl. Comput. Harmon. Anal. 35 (2013) 433–451

Contents lists available at SciVerse ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Local behavior of sparse analysis regularization: Applications to risk estimation

Samuel Vaiter [a], Charles-Alban Deledalle [a], Gabriel Peyré [a,*], Charles Dossal [b], Jalal Fadili [c]

[a] CEREMADE, CNRS, Université Paris-Dauphine, Place du Maréchal De Lattre De Tassigny, 75775 Paris Cedex 16, France
[b] IMB, Université Bordeaux 1, 351, Cours de la Libération, 33405 Talence Cedex, France
[c] GREYC, CNRS-ENSICAEN-Université de Caen, 6, Bd du Maréchal Juin, 14050 Caen Cedex, France

ARTICLE INFO

Article history:
Received 17 April 2012
Revised 10 October 2012
Accepted 29 November 2012
Available online 5 December 2012
Communicated by Dominique Picard

Keywords:
Sparsity
Analysis regularization
Inverse problems
ℓ^1 minimization
Local behavior
Degrees of freedom
SURE
GSURE
Unbiased risk estimation

ABSTRACT

In this paper, we aim at recovering an unknown signal x_0 from noisy measurements $y = \Phi x_0 + w$, where Φ is an ill-conditioned or singular linear operator and w accounts for some noise. To regularize such an ill-posed inverse problem, we impose an analysis sparsity prior. More precisely, the recovery is cast as a convex optimization program where the objective is the sum of a quadratic data fidelity term and a regularization term formed of the ℓ^1-norm of the correlations between the sought after signal and atoms in a given (generally overcomplete) dictionary. The ℓ^1-sparsity analysis prior is weighted by a regularization parameter $\lambda > 0$. In this paper, we prove that any minimizer of this problem is a piecewise-affine function of the observations y and the regularization parameter λ. As a byproduct, we exploit these properties to get an objectively guided choice of λ. In particular, we develop an extension of the Generalized Stein Unbiased Risk Estimator (GSURE) and show that it is an unbiased and reliable estimator of an appropriately defined risk. The latter encompasses special cases such as the prediction risk, the projection risk and the estimation risk. We apply these risk estimators to the special case of ℓ^1-sparsity analysis regularization. We also discuss implementation issues and propose fast algorithms to solve the ℓ^1-analysis minimization problem and to compute the associated GSURE. We finally illustrate the applicability of our framework to parameter(s) selection on several imaging problems.

1. Introduction

1.1. Regularization of linear inverse problems

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

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

where $w \in \mathbb{R}^Q$ is the noise component and the mapping $\Phi : \mathbb{R}^N \to \mathbb{R}^Q$ is a known linear operator which generally models an acquisition process that entails loss of information so that $Q \leqslant N$. Even when $Q = N$, Φ is typically ill-conditioned or even rank-deficient. In image processing, typical applications covered by the above degradation model are entry-wise

* Corresponding author.
E-mail addresses: samuel.vaiter@ceremade.dauphine.fr (S. Vaiter), deledalle@ceremade.dauphine.fr (C.-A. Deledalle), gabriel.peyre@ceremade.dauphine.fr (G. Peyré), charles.dossal@math.u-bordeaux1.fr (C. Dossal), Jalal.Fadili@greyc.ensicaen.fr (J. Fadili).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2012.11.006

masking (inpainting), convolution (acquisition blur), Radon transform (tomography) or a random sensing matrix (compressed sensing).

Solving for an accurate approximation of x_0 from the system (1) is generally ill-posed [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 object x_0. This prior information accounts for the smoothness of the solution and can range from uniform smoothness assumption to more complex geometrical priors.

Regularization is a popular way to impose such a prior, hence making the search for solutions feasible. The general variational problem we consider can be stated as

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

where F is the so-called data fidelity term, R is an appropriate regularization term that encodes the prior on the sought after signal, and $\lambda > 0$ a regularization parameter. This parameter balances between the amount of allowed noise and the regularity dictated by R. In this paper, we consider a quadratic data fidelity term taking the form

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

If it were to be interpreted in Bayesian language, this data fidelity would amount to assuming that the noise is white Gaussian.

The popular Thikonov class of regularizations corresponds to quadratic forms $R(x) = \langle x, Kx\rangle$, where K is a symmetric semidefinite positive kernel. This typically induces some kind of uniform smoothness on the recovered vector. To capture the more intricate geometrical complexity of image structures, non-quadratic priors are required, among which sparse regularization through the ℓ^1-norm has received a considerable interest in the recent years. This non-smooth regularization is at the heart of this paper.

1.2. Sparse ℓ^1-analysis regularization

We call a dictionary $D = (d_i)_{i=1}^P$ a collection of P atoms $d_i \in \mathbb{R}^N$. The dictionary may be redundant in $\mathbb{R}^N$, in which case $P > N$ and Φ if surjective if it has full row rank. D can also be viewed as a linear mapping from $\mathbb{R}^P$ to $\mathbb{R}^N$ which is used to *synthesize* a signal $x \in \operatorname{Span}(D) \subseteq \mathbb{R}^N$ 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 ℓ^1-analysis regularization in a dictionary D corresponds to using $R = R_A$ in (2) where

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

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

$$x_\lambda^\star(y) \in \operatorname*{Arg\,min}_{x\in\mathbb{R}^N} \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda\left\|D^* x\right\|_1. \tag{$P_\lambda(y)$}$$

Since the objective function in ($P_\lambda(y)$) is proper (i.e. not infinite everywhere), continuous and convex, the set of (global) minimizers of ($P_\lambda(y)$) is non-empty and compact if, and only if,

$$\operatorname{Ker}\Phi \cap \operatorname{Ker} D^* = \{0\}. \tag{H_0}$$

All throughout this paper, we suppose that this condition holds.

The most popular ℓ^1-analysis sparsity-promoting regularization is the total variation, which was first introduced for denoising (in a continuous setting) in [2]. In a discrete setting, it corresponds to taking D^* as a finite difference discretization of the gradient operator. The corresponding prior R_A favors piecewise constant signals and images. A comprehensive review of total variation regularization can be found in [3].

The theoretical properties of total variation regularization have been previously investigated. A distinctive feature of this regularization is its tendency to yield a staircasing effect, where discontinuities not present in the original data might be artificially created by the regularization. This effect has been studied by Nikolova in the discrete case in a series of papers, see e.g. [4], and by Ring in [5] in the continuous setting. The stability of the discontinuity set of the solution of the 2-D continuous total variation denoising is studied in [6].

When D is the standard basis, i.e. $D = \mathrm{Id}$, the analysis sparsity regularization R_A specializes to the so-called synthesis regularization. The corresponding variational problem ($P_\lambda(y)$) is referred to as the Lasso problem in the statistics community [7] and Basis-Pursuit DeNoising (BPDN) in the signal processing community [8]. Despite synthesis and analysis regularizations both minimize objective functions involving the ℓ^1-norm, the properties of their respective minimizers may differ significantly. Some insights on the relation and distinction between analysis and synthesis-based sparsity regularizations were first given in [9]. When D is orthogonal, and more generally when D is square and invertible, analysis and synthesis regularizations are equivalent in the sense that the set of minimizers of one problem can be retrieved from that of an equivalent form of the other through a bijective change of variable. However, when D is redundant, synthesis and analysis regularizations depart significantly.

While the theoretical guarantees of synthesis ℓ^1-regularization have been extensively studied, the analysis case remains much less investigated [10–13].

1.3. Geometrical insights into ℓ^1-analysis regularization

In the synthesis prior, sparsity of a vector $\alpha \in \mathbb{R}^P$ is measured in terms of its ℓ^0 pseudo-norm, or equivalently the cardinality of its support $\mathrm{supp}(\alpha)$, i.e.

$$\|\alpha\|_0 = \big|\mathrm{supp}(\alpha)\big| = \big|\big\{i \in \{1,\dots,P\} \setminus \alpha_i \neq 0\big\}\big|.$$

In the analysis prior, the sparsity is measured on the correlation vector D^*x. It then appears natural to keep track of the support of D^*x. To fix terminology, we define this support and its complement.

Definition 1. The *D-support* I (respectively *D-cosupport* J) of a vector $x \in \mathbb{R}^N$ is defined as $I = \mathrm{supp}(D^*x)$ (respectively $J = I^c = \{1,\dots,P\} \setminus I$).

A vector x with a D-cosupport J is then such that the correlations between this vector and the columns of D_J are zero. This is equivalent to saying that x lives in a subspace $\mathcal{G}_J$ defined as follows.

Definition 2. Given J a subset of $\{1,\dots,P\}$, the *cospace* $\mathcal{G}_J$ is defined as

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

It was shown in [13] that the subspace $\mathcal{G}_J$ plays a pivotal role in robustness and identifiability guarantees of $(P_\lambda(y))$.

In fact, the subspaces $\mathcal{G}_J$ carry all necessary information to characterize signal models of sparse analysis type. More precisely, vectors of cosparsity $k = |J|$ live in a *union of subspaces*

$$\Theta_k = \big\{\mathcal{G}_J \setminus J \subseteq \{1,\dots,P\} \text{ and } \dim \mathcal{G}_J = k\big\},$$

and the signal space $\mathbb{R}^N$ can be decomposed as $\mathbb{R}^N = \bigcup_{k\in\{0,\dots,N\}} \Theta_k$. This model has been introduced in [12] under the name *cosparse model*.

For synthesis sparsity, i.e. $D = \mathrm{Id}$, Θ_k are nothing but the set of axis-aligned subspaces of dimension k. For the 1-D total variation prior, where D corresponds to finite forward differences, Θ_k is the set of piecewise constant signals with $k-1$ steps. A few other examples of subspaces Θ_k, including those corresponding to translation invariant wavelets, are discussed in [12]. More general union of subspaces models have been introduced in sampling theory to model various types of non-linear signal ensembles, see e.g. [14].

1.4. Local behavior of minimizers

Local variations and sensitivity/perturbation analysis of problems in the form of (2) is an important topic in optimization and optimal control. Comprehensive monograph on the subject are [15,16]. 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).

In the synthesis case ($D = \mathrm{Id}$) with $Q > N$, the work of [17,18] showed that, for a fixed y, the mapping $\lambda \mapsto x^\star_\lambda(y)$ is piecewise affine, i.e. the solution path is polygonal. Further, they characterized changes in the solution $x^\star_\lambda(y)$ at vertices of this path. Based on these observations, they presented the homotopy algorithm, which follows the solution path by jumping from vertex to vertex of this polygonal path. This idea was extended to the underdetermined case in [19,20]. A homotopy-type scheme was proposed in [21] for sparse ℓ^1-analysis regularization in the overdetermined case ($Q > N$). We will discuss the latter work in more detail in Section 4.

1.5. Risk estimation and parameter selection

This paper also addresses unbiased estimation of the ℓ^2-risk when recovering a vector $x_0 \in \mathbb{R}^N$ from the measurements y in (1), e.g. by solving (2), under the assumption that w is white Gaussian noise. A central concept for risk estimation is that of the degrees of freedom (DOF). Let $\hat{x}_\theta(y)$ be an estimator of x_0 from (1), parameterized by some parameters θ. The DOF of such an estimator was defined by Efron [22] as

$$df_\theta = \sum_{i=1}^{Q} \frac{\mathrm{cov}_w(y_i, (\Phi\hat{x}_\theta(y))_i)}{\sigma^2}.$$

The DOF is generally used to quantify the complexity of a statistical modeling procedure. It plays a central role in many model validation and selection criteria, e.g. Mallows' C_p (Mallows [23]), AIC (Akaike information criterion [24]), BIC (Bayesian information criterion [25]), GCV (generalized cross-validation [26]) or SURE (Stein Unbiased Risk Estimator [27]). In the spirit of the SURE theory, a good unbiased estimator of the DOF is sufficient to provide an unbiased estimate of the ℓ^2-risk in reconstructing Φx_0, i.e. the *prediction* risk $\mathbb{E}_w(\|\Phi\hat{x}_\theta(y) - \Phi x_0\|_2^2)$. For instance, the SURE is given by

$$\mathrm{SURE}\big(\hat{x}_\theta(y)\big) = \big\|y - \Phi\hat{x}_\theta(y)\big\|_2^2 - Q\sigma^2 + 2\sigma^2\widehat{df}_\theta(y) \tag{5}$$

with

$$\widehat{df}_\theta(y) = \mathrm{tr}\left(\frac{\partial \Phi\hat{x}_\theta(y)}{\partial y}\right),$$

where $\mathbb{E}_w(\widehat{df}_\theta(y)) = df_\theta$, $\frac{\partial \Phi\hat{x}_\theta(y)}{\partial y}$ is the Jacobian matrix of the vector function $y \mapsto \Phi\hat{x}_\theta(y)$ and tr is the trace operator.

The SURE depends solely on y, without prior knowledge of x_0. This can prove very useful as a basis to automatically choose the parameters θ of the estimator, e.g. $\theta = \lambda$ when solving (2), or the smoothing parameters in families of linear estimates [28] such as for ridge regression or smoothing splines. In some settings, it has been shown that it offers better accuracy than GCV and related non-parametric selection techniques [29]. However, compared to GCV, the SURE requires the knowledge of the noise variance σ^2.

The SURE has been widely used in the statistics and signal processing communities as a principled and efficient way for parameter selection with a variety of non-linear estimators. For instance, it was used for wavelet denoising [30–32], wavelet shrinkage for a class of linear inverse problems [33] and non-local filtering [34–36].

For general linear inverse problems, the estimator of the prediction risk and the parameter(s) minimizing it can depart substantially from those corresponding to the *estimation* risk $\mathbb{E}_w(\|\hat{x}_\theta(y) - x_0\|_2^2)$ [37]. To circumvent this difficulty, in [38], the authors attempted to approximate the estimation risk by relying on a regularized version of the inverse of Φ. However, in general, either Φ should have a trivial kernel or, otherwise, x_0 should live outside to $\ker(\Phi)$ to guarantee the existence of an unbiased estimator of the estimation risk [39].

In [40], a generalized SURE (GSURE) has been developed for noise models within the multivariate canonical exponential family. This allows one to derive an unbiased estimator of the risk on a projected version of $\hat{x}_\theta(y)$, which in turn covers the case where Φ has a non-trivial kernel and a part of x_0 is in it. Indeed, in the latter scenario, the GSURE can at best estimate the *projection* risk $\mathbb{E}_w(\|\Pi\hat{x}_\theta(y) - \Pi x_0\|_2^2)$ where Π is the orthogonal projector on $\ker(\Phi)^\perp$.

1.6. Contributions

This paper describes the following contributions:

1. *Local affine parameterization*: we show that any solution $x_\lambda^\star(y)$ of $(P_\lambda(y))$ is a piecewise-affine function of (y, λ). Furthermore, for fixed λ, and for y outside a set of Lebesgue measure zero, the prediction $\mu_\lambda^\star(y)$ locally varies along a constant subspace. This is a distinctly novel contribution which generalizes previously known results (see Section 4.1 for a detailed discussion). It also forms the cornerstone of unbiased estimation of the DOF.
2. *GSURE*: we derive a unifying framework to compute unbiased estimates of several risks in ℓ^2 sense, for estimators of x_0 from y as observed in (1) when w is a white Gaussian noise. This framework encompasses for instance the prediction, the projection and the estimation risks (see Section 4.3 for a discussion to related work).
3. *ℓ^1-analysis unbiased risk estimation*: combining the results from the previous two contributions, we derive a closed-form expression of an unbiased estimator of the DOF for $(P_\lambda(y))$, whence we deduce GSURE estimates of the different risks.
4. *Numerical computation of GSURE*: we also address in detail numerical issues that rise when implementing our DOF estimator and GSURE for $(P_\lambda(y))$. We show that the additional computational effort to compute the DOF estimator (hence the GSURE) from its closed-form is invested in solving simple linear systems. This turns out to be much faster than iterative approaches existing in the literature which are computationally demanding (see Section 4.4 for a detailed discussion).

1.7. Organization of the paper

The rest of the paper is organized as follows. Sections 2 and 3 describe each of our main contributions. Section 4 draws some connections with relevant previous works. Section 5 illustrates our results on some numerical examples. The proofs are deferred to Appendix A awaiting inspection by the interested reader.

1.8. Notation

We first summarize the main notations used throughout the paper. We focus on real vector spaces. The sign vector $\mathrm{sign}(\alpha)$ of $\alpha \in \mathbb{R}^P$ is

$$\forall i \in \{1, \ldots, P\}, \quad \mathrm{sign}(\alpha)_i = \begin{cases} +1 & \text{if } \alpha_i > 0, \\ 0 & \text{if } \alpha_i = 0, \\ -1 & \text{if } \alpha_i < 0. \end{cases}$$

Its support is

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

For a subset $I \subset E$, $|I|$ will denote its cardinality, and $I^c = E \setminus I$ its complement.

The matrix M_J for J a subset of $\{1, \dots, P\}$ is the submatrix whose columns are indexed by J. Similarly, the vector s_J is the restriction of s to the entries of s indexed by J.

tr and div are respectively the trace and divergence operators. The matrix Id is the identity matrix, where the underlying space will be clear from the context. For any matrix M, M^+ is its Moore–Penrose pseudoinverse and M^* is its adjoint.

2. Perturbation theory of ℓ^1-analysis regularization

Throughout this section, it is important to point out that we only require that the noise vector $w \in \mathbb{R}^Q$ to be bounded. The fact that it could be deterministic or random is irrelevant here.

2.1. Local affine parameterization

Our first contribution derives a local affine parameterization of minimizers of $(P_\lambda(y))$ as functions of $(y, \lambda) \in \mathbb{R}^Q \times \mathbb{R}_+$. To develop our theory, the invertibility of Φ on $\mathcal{G}_J$ will play a vital role. For this, we need to assume that

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

To intuitively understand the importance of this assumption, think of the ideal case where one wants to estimate a D-sparse signal x_0 from $y = \Phi x_0 + w$, whose D-cosupport J is assumed to be known. This can be achieved by solving a least-squares problem. The latter has a unique solution if (H_J) holds.

Of course, J is not known in general, and one may legitimately ask whether (H_J) is fulfilled for some solution of $(P_\lambda(y))$. We will provide an affirmative answer to this question in Theorem 2(ii), i.e. there always exists a solution of $(P_\lambda(y))$ such that (H_J) holds.

With assumption (H_J) at hand, we now define the following matrix whose role will be clarified shortly.

Definition 3. Let J be a D-cosupport. Suppose that (H_J) holds. We define the matrix $\Gamma^{[J]}$ 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 $\mathcal{G}_J$.

Observe that the action of $\Gamma^{[J]}$ could be rewritten as an optimization problem

$$\Gamma^{[J]}u = \underset{D_J^* x = 0}{\arg\min} \frac{1}{2}\|\Phi x\|^2 - \langle x, u\rangle.$$

Let us now turn to sensitivity of the minimizers $x_\lambda^\star(y)$ of $(P_\lambda(y))$ to perturbations of (y, λ). More precisely, our aim is to study properties, including continuity and differentiability, of $x_\lambda^\star(y)$ and $\Phi x_\lambda^\star(y)$ as functions of y and λ. Toward this end, we will exploit the fact that $x_\lambda^\star(y)$ obeys an implicit equation given in Lemma 2 (see Appendix A.2). But as optimal solutions turn out to be not everywhere differentiable (change of the D-support and thus of the cospace), we will concentrate on a local analysis where (y, λ) vary in a small neighborhood that typically avoids non-differentiability to occur. This is exactly the reason why we introduce the transition space $\mathcal{H}$ defined below. It corresponds to the set of observation vectors y and regularization parameters λ where the cospace $\mathcal{G}_J$ of any solution of $(P_\lambda(y))$ is not stable with respect to small perturbations of (y, λ).

Definition 4. The *transition space* $\mathcal{H}$ is defined as

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

where

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

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

The following theorem summarizes our first sensitivity analysis result on the optimal solutions of $(P_\lambda(y))$.

Theorem 1. *Let $(y,\lambda) \notin \mathcal{H}$ and let $x^\star_\lambda(y)$ be a solution of $(P_\lambda(y))$. Let I and J be the D-support and D-cosupport of $x^\star_\lambda(y)$ and $s = \operatorname{sign}(D^* x^\star_\lambda(y))$. Suppose that (H_J) holds. For any $\bar{y} \in \mathbb{R}^Q$ and $\bar{\lambda} \in \mathbb{R}_+$, define*

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

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

An immediate consequence of this theorem is that, for a fixed $y \in \mathbb{R}^Q$, if $(P_\lambda(y))$ admits a unique solution $x^\star_\lambda(y)$ for each λ, then $\{x^\star_\lambda(y):\ \lambda \in \mathbb{R}_+\}$ identifies a polygonal solution path. As we move along the solution path, the cospace is piecewise constant as a function of λ, changing only at critical values corresponding to the vertices on the polygonal path.

2.2. Local variations of the prediction

We now turn to quantifying explicitly the local variations of the prediction $\mu^\star_\lambda(y) = \Phi x^\star_\lambda(y)$ with respect to the observation y. First, it is not difficult to see that even if $(P_\lambda(y))$ admits several solutions, all of them share the same image under Φ; see Lemma 4 for a formal proof of this assertion. This allows to denote without ambiguity $\mu^\star_\lambda(y)$ as a single-valued mapping. Before stating our second sensitivity analysis result, we need to define the restriction to $\mathbb{R}^Q$ of the transition space $\mathcal{H}$.

Definition 5. Let $\lambda \in \mathbb{R}^*_+$. The *$\lambda$-restricted transition space* is

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

Theorem 2. *Fix $\lambda \in \mathbb{R}^*_+$. Then,*

(i) *the λ-restricted transition space $\mathcal{H}_{\cdot,\lambda}$ is of Lebesgue measure zero*;
(ii) *for $y \notin \mathcal{H}_{\cdot,\lambda}$, there exists $x^\star_\lambda(y)$ a solution of $(P_\lambda(y))$ with a D-cosupport J that obeys (H_J)*;
(iii) *the mapping $y \mapsto \mu^\star_\lambda(y)$ is of class C^∞ on $\mathbb{R}^Q \setminus \mathcal{H}_{\cdot,\lambda}$ (a set of full Lebesgue measure), and*

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

where J is such that (H_J) holds.

3. Generalized Stein unbiased risk estimator

Throughout this section, for our statements to be statistically meaningful, the noise is assumed to be white Gaussian, $w \sim \mathcal{N}(0, \sigma^2 \operatorname{Id}_Q)$ of bounded variance σ^2.

3.1. GSURE for an arbitrary estimator

We first consider an arbitrary estimator $\hat{x}_\theta(y)$ with parameters θ such that $\hat{\mu}_\theta(y) = \Phi \hat{x}_\theta(y)$ is a single-valued mapping. We similarly write $\mu_0 = \Phi x_0$. Of course the results described shortly will apply when the estimator is taken as any minimizer of $(P_\lambda(y))$, in which case $\theta = \lambda$.

We here develop an extended version of GSURE that unbiasedly estimates the risk of reconstructing $A\mu_0$ with an arbitrary matrix $A \in \mathbb{R}^{M \times Q}$. This allows us to cover in a unified framework unbiased estimation of several classical risks including the prediction risk (with $A = \operatorname{Id}$), the projection risk when Φ is rank-deficient (with $A = \Phi^*(\Phi\Phi^*)^+$), and the estimation risk when Φ has full rank (with $A = \Phi^+ = (\Phi^*\Phi)^{-1}\Phi^*$). A quantity that will enter into play in the risk of estimating $A\mu_0$ is the degrees of freedom defined as

$$df^A_\theta = \sum_{i=1}^{Q} \frac{\operatorname{cov}_w((Ay)_i, (A\hat{\mu}_\theta(y))_i)}{\sigma^2}.$$

Definition 6. Let $A \in \mathbb{R}^{M \times Q}$. We define the Generalized Stein Unbiased Risk Estimate (GSURE) associated to A as

$$\operatorname{GSURE}^A\big(\hat{x}_\theta(y)\big) = \big\| A\big(y - \hat{\mu}_\theta(y)\big) \big\|_2^2 - \sigma^2 \operatorname{tr}\big(A^*A\big) + 2\sigma^2 \widehat{df}^A_\theta(y),$$

where

$$\widehat{df}^A_\theta(y) = \operatorname{tr}\left(A \frac{\partial \hat{\mu}_\theta(y)}{\partial y} A^* \right).$$

Unbiasedness of the GSURE The next result shows that $\operatorname{GSURE}^A(\hat{x}_\theta(y))$ is an unbiased estimator of an appropriate ℓ^2-risk, and $\widehat{df}^A_\theta(y)$ is an unbiased estimator of df^A_θ.

Theorem 3. *Let $A \in \mathbb{R}^{M\times Q}$. Suppose that $y \mapsto \hat{\mu}_\theta(y)$ is weakly differentiable, so that its divergence is well-defined in the weak sense. If $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$, then*

$$\mathbb{E}_w \,\mathrm{GSURE}^A\big(\hat{x}_\theta(y)\big) = \mathbb{E}_w\big(\big\|A\mu_0 - A\hat{\mu}_\theta(y)\big\|_2^2\big) \quad \text{and} \quad \mathbb{E}_w \widehat{df}_\theta^A(y) = df_\theta^A.$$

Remark 1. Theorem 3 can be straightforwardly adapted to deal with any white Gaussian noise with a non-singular covariance matrix Σ. It is sufficient to consider the change of variable $y \mapsto \Sigma^{-1/2}y$ and $\Phi \mapsto \Sigma^{-1/2}\Phi$. This is in the same vein as [40].

All estimators of the form GSURE^B with B such that $B\Phi = A\Phi$ share the same expectation given by Theorem 3. Hence, there are several ways to estimate the risk in reconstructing $A\mu_0$. For the estimation of the prediction, projection and estimation risks, we now give the corresponding expressions and associated estimators (with subscript notations) as direct consequences of Theorem 3:

- $A = \mathrm{Id}$: in which case $\mathrm{GSURE}^{\mathrm{Id}}$ becomes

  $$\mathrm{GSURE}_\Phi\big(\hat{x}_\theta(y)\big) = \big\|y - \hat{\mu}_\theta(y)\big\|_2^2 - Q\sigma^2 + 2\sigma^2 \,\mathrm{tr}\left(\frac{\partial \hat{\mu}_\theta(y)}{\partial y}\right)$$

  which provides an unbiased estimate of the prediction risk

  $$\mathrm{Risk}_\Phi(x_0) = \mathbb{E}_w\big\|\Phi\hat{x}_\theta(y) - \Phi x_0\big\|_2^2.$$

  This coincides with the classical SURE defined in (5).
- $A = \Phi^*(\Phi\Phi^*)^+$: when Φ is rank-deficient, $\Pi = \Phi^*(\Phi\Phi^*)^+\Phi$ is the orthogonal projector on $\ker(\Phi)^\perp = \mathrm{Im}(\Phi^*)$. Denoting $x_{\mathrm{ML}}(y) = \Phi^*(\Phi\Phi^*)^+ y$ the maximum likelihood estimator (MLE), $\mathrm{GSURE}^{\Phi^*(\Phi\Phi^*)^+}$ becomes

  $$\mathrm{GSURE}_\Pi\big(\hat{x}_\theta(y)\big) = \big\|x_{\mathrm{ML}}(y) - \Pi\hat{x}_\theta(y)\big\|_2^2 - \sigma^2 \,\mathrm{tr}\big(\big(\Phi\Phi^*\big)^+\big) + 2\sigma^2 \,\mathrm{tr}\left(\big(\Phi\Phi^*\big)^+ \frac{\partial \hat{\mu}_\theta(y)}{\partial y}\right).$$

  It provides an unbiased estimate of the projection risk

  $$\mathrm{Risk}_\Pi(x_0) = \mathbb{E}_w\big\|\Pi\hat{x}_\theta(y) - \Pi x_0\big\|_2^2.$$

  If Φ is the synthesis operator of a Parseval tight frame, i.e. $\Phi\Phi^* = \mathrm{Id}$, the projection risk coincides with the prediction risk and so do the corresponding GSURE estimates

  $$\mathrm{Risk}_\Pi(x_0) = \mathrm{Risk}_\Phi(x_0) \quad \text{and} \quad \mathrm{GSURE}_\Pi\big(\hat{x}_\theta(y)\big) = \mathrm{GSURE}_\Phi\big(\hat{x}_\theta(y)\big).$$

  It is also worth noting that if $\hat{x}_\theta(y)$ never lies in $\ker(\Phi)$, then $\mathrm{Risk}_\Pi(x_0)$ coincides with the estimation risk up to the additive constant $\|(\mathrm{Id} - \Pi)x_0\|_2^2$.
- $A = (\Phi^*\Phi)^{-1}\Phi^*$: in this case Φ has full rank, and the mapping $y \mapsto \hat{x}_\theta(y)$ is single-valued and weakly differentiable. The maximum likelihood estimator is now $x_{\mathrm{ML}}(y) = (\Phi^*\Phi)^{-1}\Phi^* y$, and $\mathrm{GSURE}^{(\Phi^*\Phi)^{-1}\Phi^*}$ takes the form

  $$\mathrm{GSURE}_{\mathrm{Id}}\big(\hat{x}_\theta(y)\big) = \big\|x_{\mathrm{ML}}(y) - \hat{x}_\theta(y)\big\|_2^2 - \sigma^2 \,\mathrm{tr}\big(\big(\Phi^*\Phi\big)^{-1}\big) + 2\sigma^2 \,\mathrm{tr}\left(\Phi\big(\Phi^*\Phi\big)^{-1}\frac{\partial \hat{x}_\theta(y)}{\partial y}\right).$$

  This is an unbiased estimator of the estimation risk given by

  $$\mathrm{Risk}_{\mathrm{Id}}(x_0) = \mathbb{E}_w\big\|\hat{x}_\theta(y) - x_0\big\|_2^2.$$

Reliability of the GSURE We now assess the reliability of the GSURE by computing the expected squared-error between $\mathrm{GSURE}^A(\hat{x}_\theta(y))$ and the true squared-error on $A\mu_0$

$$\mathrm{SE}^A\big(\hat{x}_\theta(y)\big) = \big\|A\mu_0 - A\hat{\mu}_\theta(y)\big\|_2^2.$$

Theorem 4. *Under the assumptions of Theorem 3, we have*

$$\begin{aligned}
&\mathbb{E}_w\big[\big(\mathrm{GSURE}^A\big(\hat{x}_\theta(y)\big) - \mathrm{SE}^A\big(\hat{x}_\theta(y)\big)\big)^2\big] \\
&\quad = 2\sigma^4 \,\mathrm{tr}\big[\big(A^*A\big)^2\big] + 4\sigma^2\mathbb{E}_w\big\|A^*A\big(\mu_0 - \hat{\mu}_\theta(y)\big)\big\|_2^2 \\
&\qquad - 4\sigma^4\mathbb{E}_w\left(\mathrm{tr}\left[A\frac{\partial \hat{\mu}_\theta(y)}{\partial y}A^*A\left(2\,\mathrm{Id} - \frac{\partial \hat{\mu}_\theta(y)}{\partial y}\right)A^*\right]\right).
\end{aligned}$$

3.2. GSURE for ℓ^1-analysis regularization

We now specialize the previous results to the case where the estimator $\hat{x}_\theta(y)$ is a solution of $(P_\lambda(y))$; i.e. $\hat{x}_\theta(y) = x^\star_\lambda(y)$ and $\hat{\mu}_\theta(y) = \mu^\star_\lambda(y)$. For notational clarity and to highlight the dependency of $\dim(\mathcal{G}_J)$ on y, for $y \notin \mathcal{H}_{\cdot,\lambda}$, we write $d(y) = \dim(\mathcal{G}_J)$ where J is the D-cosupport of any solution $x^\star_\lambda(y)$ such that (H_J) holds. We then obtain the following corollary as a consequence of Theorems 2 and 3.

Corollary 1. *Let $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$. Then $\mu^\star_\lambda(y)$ is weakly differentiable and*

$$\begin{aligned}
\mathrm{GSURE}_\Phi\big(x^\star_\lambda(y)\big) &= \big\|y - \mu^\star_\lambda(y)\big\|_2^2 - Q\sigma^2 + 2\sigma^2 d(y),\\
\mathrm{GSURE}_\Pi\big(x^\star_\lambda(y)\big) &= \big\|x_{\mathrm{ML}}(y) - \Pi x^\star_\lambda(y)\big\|_2^2 - \sigma^2 \operatorname{tr}\big(\big(\Phi\Phi^*\big)^+\big) + 2\sigma^2 \operatorname{tr}\big(\Pi \Gamma^{[J]}\big),\\
\mathrm{GSURE}_{\mathrm{Id}}\big(x^\star_\lambda(y)\big) &= \big\|x_{\mathrm{ML}}(y) - x^\star_\lambda(y)\big\|_2^2 - \sigma^2 \operatorname{tr}\big(\big(\Phi^*\Phi\big)^{-1}\big) + 2\sigma^2 \operatorname{tr}\big(\Gamma^{[J]}\big).
\end{aligned}$$

Moreover, $d(y)$ is an unbiased estimator of the DOF of $(P_\lambda(y))$, i.e.

$$df_\lambda = df^{\mathrm{Id}}_\lambda = \mathbb{E}_w d(y).$$

In particular, this result states that $\dim(\mathcal{G}_J)$ is an unbiased estimator of the DOF of $(P_\lambda(y))$ response without requiring any assumption to ensure uniqueness of $x^\star_\lambda(y)$. This DOF estimator formula is valid everywhere except on a set of (Lebesgue) measure zero.

Building upon Theorems 2 and 4, we derive the relative reliability of the GSURE for $(P_\lambda(y))$, and show that it decays with the number of measurements at the rate $O(1/Q)$.

Corollary 2. *Let $A \in \mathbb{R}^{M\times Q}$ and $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$. Then*

$$\mathbb{E}_w\left[\left(\frac{\mathrm{GSURE}^A(x^\star_\lambda(y)) - \mathrm{SE}^A(x^\star_\lambda(y))}{Q\sigma^2}\right)^2\right] = O\left(\frac{\|A\|^4}{Q}\right),$$

where $\|A\|$ is the spectral norm of A. In particular, if $\|A\|$ is independent of Q, the decay rate of the relative reliability is $O(1/Q)$.

3.3. Numerical considerations

The remaining obstacle faced when implementing the GSURE formulae of Corollary 1 is to compute the divergence term, i.e. the last trace term as given by $\widehat{df}^A_\lambda(y) = \operatorname{tr}(A\Phi\Gamma^{[J]}\Phi^*A^*)$ (see Definition 6). However, for large-scale data as in image and signal processing, the computational storage required for the matrix in the argument of the trace would be prohibitive. Additionally, computing $\Gamma^{[J]}$ can only be reasonably afforded for small data size. Fortunately, the structure of $\widehat{df}^A_\lambda(y)$ and the definition of $\Gamma^{[J]}$ allows to derive an efficient and principled way to compute the trace term. This is formalized in the next result.

Proposition 1. *One has*

$$\widehat{df}^A_\lambda(y) = \mathbb{E}_Z\big(\langle \nu(Z), \Phi^*A^*AZ\rangle\big) \tag{8}$$

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

$$\frac{1}{k}\sum_{i=1}^{k}\langle \nu(z_i), \Phi^*A^*Az_i\rangle \xrightarrow{\mathrm{WLLN}} \widehat{df}^A_\lambda(y), \tag{10}$$

for k realizations z_i of Z, where WLLN stands for the Weak Law of Large Numbers. Consequently, the computational bulk of computing an estimate of $\widehat{df}^A_\lambda(y)$ is invested in solving for each $\nu(z_i)$ the symmetric linear system (9) using e.g. a conjugate gradient solver.

4. Relation to other works

4.1. Local variations

The local behavior of $x_\lambda^\star(y)$ as a function of λ is already known in the ℓ^1-synthesis case, both for the case where Φ is full rank [17,18], and $Q < N$ [20]. Our local affine parameterization in Theorem 1 generalizes these results to the analysis case regardless of the number of measurements. Our result also goes beyond the work of [21] which investigates the overdetermined case with an ℓ^1-analysis regularization and develops a homotopy algorithm.

4.2. Degrees of freedom

In the synthesis overdetermined case with full rank Φ, [41] showed that the number of nonzero coefficients is an unbiased estimate for the degrees of freedom of $(P_\lambda(y))$. This was generalized to an arbitrary Φ in [42]. Corollary 1 encompasses these results as special cases by taking $D = \mathrm{Id}$.

For the ℓ^1-analysis regularization with full rank Φ, Tibshirani and Taylor [21] showed that $df_\lambda = \mathbb{E}_w \dim(\mathcal{G}_J)$, where J is the D-cosupport of the unique solution to $(P_\lambda(y))$. This is exactly the assertion of Corollary 1, since (H_J) is in force when $\operatorname{rank}(\Phi) = N$.

While a first version of this paper was submitted, it came to our attention that Tibshirani and Taylor [43, Theorem 3] recently and independently developed an unbiased estimator of the DOF for $(P_\lambda(y))$ that covers the case where $Q < N$. More precisely, they showed that $\dim(\Phi(\mathcal{G}_J))$ is an unbiased estimator of $df(\lambda)$, where J is the D-cosupport of *any* solution to $(P_\lambda(y))$. This coincides with Corollary 1 when J satisfies (H_J). Their proof however differs from ours, and in particular, it does not study directly the local behavior of $x_\lambda^\star(y)$ as a function of y or λ (Theorem 1).

4.3. Generalized Stein unbiased risk estimator

In [40], the author derived expressions equivalent to GSURE_Π and $\mathrm{GSURE}_{\mathrm{Id}}$ up to a constant which does not depend on the estimator. However, her expressions were developed separately, whereas we have shown that these GSURE estimates originate from a general result stated in Theorem 3. Another distinction between our work and [40] lies in the assumptions imposed. The author [40] supposes $\hat{x}_\theta(y)$ to be a weakly differentiable function of $\Phi^* y/\sigma^2$. In contrast, we just require that the prediction $y \mapsto \hat{\mu}_\theta(y)$ (a single-valued map) is weakly differentiable, as classically assumed in the SURE theory.

Indeed, let $u = \Phi^* y/\sigma^2$, and define $\hat{x}_\theta(y) = z_\theta^\star(u)$. Assume that $u \mapsto z_\theta^\star(u)$ is weakly differentiable (and a fortiori a single-valued mapping).

When Φ is rank-deficient, [40] proves unbiasedness of the following estimator of the projection risk

$$\mathrm{GSURE}_\Pi^{(\mathrm{Eldar})}\big(z_\theta^\star(u)\big) = \|\Pi x_0\|_2^2 + \big\|\Pi z_\theta^\star(u)\big\|_2^2 - 2\big\langle z_\theta^\star(u), x_{\mathrm{ML}}(y)\big\rangle + 2\,\mathrm{tr}\left(\Pi \frac{\partial z_\theta^\star(u)}{\partial u}\right).$$

Since by assumption $\frac{\partial \Phi z_\theta^\star(u)}{\partial u} = \Phi \frac{\partial z_\theta^\star(u)}{\partial u}$, and using the chain rule, the following holds

$$\sigma^2\,\mathrm{tr}\left(\big(\Phi\Phi^*\big)^+ \frac{\partial \hat{\mu}_\theta(y)}{\partial y}\right) = \sigma^2\,\mathrm{tr}\left(\big(\Phi\Phi^*\big)^+ \frac{\partial \Phi z_\theta^\star(u)}{\partial u}\frac{\partial u}{\partial y}\right) = \mathrm{tr}\left(\Pi \frac{\partial z_\theta^\star(u)}{\partial u}\right)$$

whence it follows that

$$\mathrm{GSURE}_\Pi\big(\hat{x}_\theta(y)\big) - \mathrm{GSURE}_\Pi^{(\mathrm{Eldar})}\big(\hat{x}_\theta(y)\big) = \big\|x_{\mathrm{ML}}(y)\big\|_2^2 - \|\Pi x_0\|_2^2 - \sigma^2\,\mathrm{tr}\big(\big(\Phi\Phi^*\big)^+\big).$$

A similar reasoning when Φ has full rank leads to

$$\mathrm{GSURE}_{\mathrm{Id}}\big(\hat{x}_\theta(y)\big) - \mathrm{GSURE}_{\mathrm{Id}}^{(\mathrm{Eldar})}\big(\hat{x}_\theta(y)\big) = \big\|x_{\mathrm{ML}}(y)\big\|_2^2 - \|x_0\|_2^2 - \sigma^2\,\mathrm{tr}\big(\big(\Phi^*\Phi\big)^{-1}\big).$$

Both our estimators and those of [40] are unbiased, but they do not have necessarily the same variance. Given that they only differ by terms that do not depend on $\hat{x}_\theta(y)$, and in particular on the parameter (here θ), selecting the latter by minimizing our GSURE expressions or those of [40] is expected to lead to the same results.

Let us finally mention that in the context of deconvolution, GSURE_Π boils down to the unbiased estimator of the projection risk obtained in [44].

4.4. Numerical computation of the GSURE

In least-squares regression regularized by a sufficiently smooth penalty term, the DOF can be estimated in closed-form [45]. However even in such simple cases, the computational load and/or storage can be prohibitive for large-scale data.

To overcome the analytical difficulty for general non-linear estimators, when no closed-form expression is available, first attempts developed bootstrap-based (asymptotically) unbiased estimators of the DOF [29]. Ye [46] and Shen and Ye [47]

proposed a data perturbation technique to approximate the DOF (and the SURE) when its closed-form expression is not available or numerically expensive to compute. For denoising, a similar Monte Carlo approach has been used in [48] where it was applied to total variation denoising, wavelet soft-thresholding, and Wiener filtering/smoothing splines.

Alternatively, an estimate can be obtained by recursively differentiating the sequence of iterates that converges to a solution of the original minimization problem. Initially, it has been proposed by [38], and then refined in [49], to compute the GSURE of sparse synthesis regularization by differentiating the sequence of iterates of the forward–backward splitting algorithm. We have recently proposed a generalization of this methodology to any proximal splitting algorithm, and exemplified it on ℓ^1-analysis regularization including the isotropic total variation regularization, and ℓ^1–ℓ^2 synthesis regularization which promotes block sparsity [50].

In our case, we have shown that the computation of a good estimator of the DOF, and therefore of GSURE^A for various risks, boils down to solving linear systems. This is much more efficient than the previous general-purpose iterative methods that are computationally expensive.

5. Numerical experiments

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 $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$, the parameter λ that minimizes the value of GSURE when solving $(P_\lambda(y))$ from $y = \Phi x_0 + w$ for various scenarios on Φ and x_0.

5.1. Computing minimizers

Denoising Although it is convex, solving problem $(P_\lambda(y))$ is rather challenging given its non-smoothness. In the case where $\Phi = \mathrm{Id}$, the objective functional of $(P_\lambda(y))$ is strictly convex, and one can compute its unique solution $x^\star_\lambda(y)$ by solving an equivalent Fenchel–Rockafellar dual problem [51]

$$x^\star_\lambda(y) = y + D\alpha^\star_\lambda(y) \quad \text{where } \alpha^\star_\lambda(y) \in \operatorname*{Arg\,min}_{\|\alpha\|_\infty \leqslant \lambda} \|y + D\alpha\|_2^2.$$

This dual problem can be solved using e.g. projected gradient descent or a multi-step accelerated version.

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 a more elaborate primal–dual splitting algorithm. We use in our numerical experiments the relaxed Arrow–Hurwicz algorithm as revitalized recently in [52]. This algorithm achieves full splitting where all operators are applied separately: the proximity operators of $g \mapsto \frac{1}{2}\|y - g\|_2^2$ and $u \mapsto \lambda\|u\|_1$ (which are known in closed-form), and the linear operators Φ and D and their adjoints. To cast $(P_\lambda(y))$ in the form required to apply this scheme, we can rewrite it as

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

Note that other algorithms could be equally applied to solve $(P_\lambda(y))$, e.g. [53–55].

5.2. Parameter selection using the GSURE

Super-resolution with total variation regularization In this example, Φ is a vertical sub-sampling operator of factor two (hence $Q/N = 0.5$). The noise level has been set such that the observed image y has a peak signal-to-noise ratio (PSNR) of 27.78 dB. We used an anisotropic total variation regularization; i.e. the sum of the ℓ^1-norms of the partial derivatives in the first and second direction (not to be confused with the isotropic total variation). Fig. 1(d) depicts the projection risk and its GSURE$_{\Pi}$ estimate obtained from (10) with $k = 1$ as a function of λ. The curves appear unimodal and coincide even with $k = 1$ and a single noise realization. Consequently, GSURE$_{\Pi}$ provides a high-quality selection of λ minimizing the projection risk. Close-up views of the central parts of the degraded, restored (using the optimal λ), and true images are shown in Fig. 1(a)–(c) for visual inspection of the restoration quality.

Compressed sensing with wavelet analysis regularization We consider in this example a compressed sensing scenario where Φ is a random partial DCT measurement matrix with an under-sampling ratio $Q/N = 0.5$. The noise is such that input image y has a PSNR set to 27.50 dB. We took D as the shift-invariant Haar wavelet dictionary with 3 scales. Again, we estimate GSURE$_{\Pi}$ with $k = 1$ in (10). The results observed on the super-resolution example are confirmed in this compressed sensing experiment both visually and qualitatively, see Fig. 2.

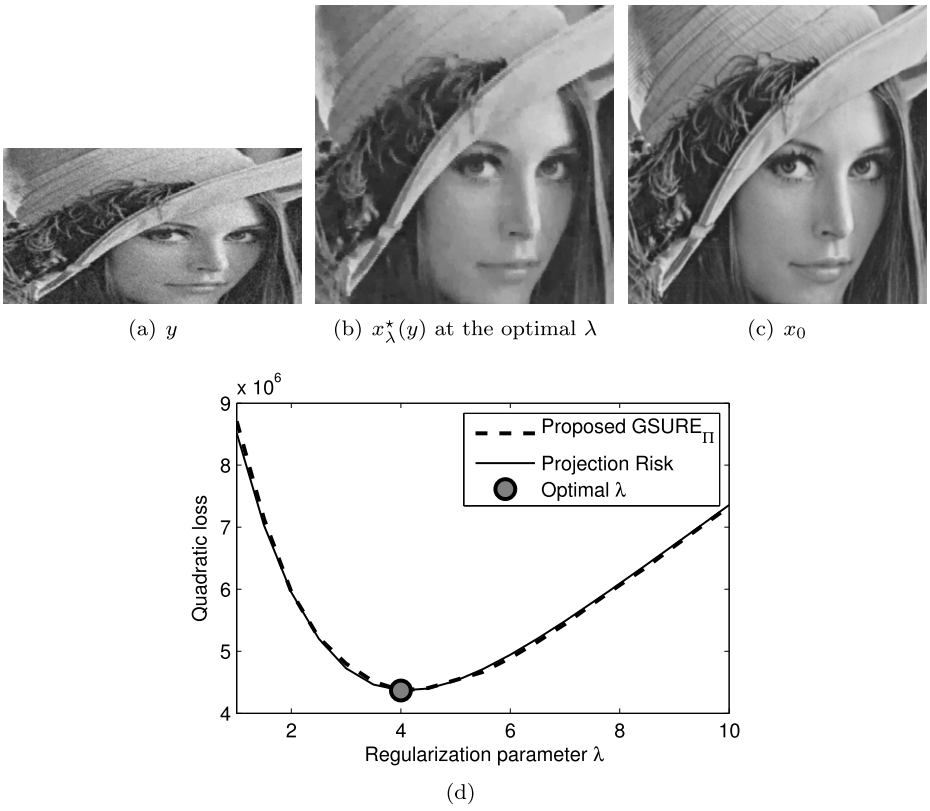


Fig. 1. Illustration of the selection of λ by minimizing GSURE_{Π} in a super-resolution problem ($Q/N = 0.5$) with anisotropic total variation regularization. (a) The observed image y. (b) A solution $x_\lambda^\star(y)$ of $(P_\lambda(y))$ at the optimal λ (the one minimizing GSURE_{Π}). (c) The underlying true image x_0. (d) Projection risk Risk_{Π} and its GSURE_{Π} estimate obtained from (10) using $k = 1$ random realization.

6. Conclusion

In this paper, we studied the local behavior of solutions to ℓ^1-analysis regularized inverse problems of the form $(P_\lambda(y))$. We proved that any minimizer $x_\lambda^\star(y)$ of $(P_\lambda(y))$ is a piecewise-affine function of the observations y and the regularization parameter λ. This local affine parameterization is completely characterized by the D-support I of $x_\lambda^\star(y)$, i.e. the set of indices of atoms in D with nonzero correlations with $x_\lambda^\star(y)$. As a byproduct, for y outside a set of zero Lebesgue measure, the first-order variations of $\Phi x_\lambda^\star(y)$ with respect to y are obtained in closed-form.

We capitalized on these results to derive a closed-form expression of an unbiased estimator of the degrees of freedom of $(P_\lambda(y))$, and to objectively and automatically choose the regularization parameter λ when the noise contaminating the observations is additive-white Gaussian. Toward this goal, a unified framework to unbiasedly estimate several risk measures is proposed through the GSURE methodology. This encompasses several special cases such as the prediction, the projection and the estimation risk. A computationally efficient algorithm is designed to compute the GSURE in the context of ℓ^1-analysis reconstruction. Illustrations on different imaging inverse problems exemplify the potential applicability of our theoretical findings.

Appendix A. Proofs

Throughout, we use the shorthand notation $\mathcal{L}_{y,\lambda}$ for the objective function in $(P_\lambda(y))$

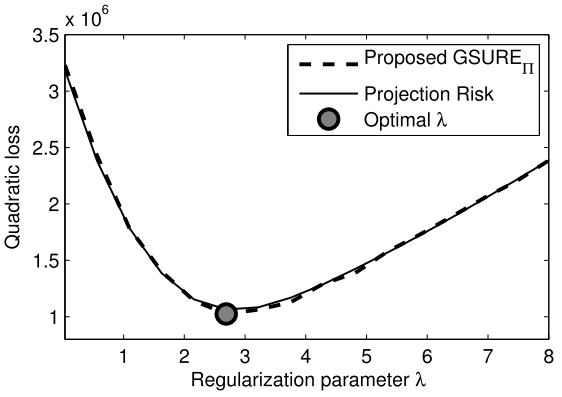
$$\mathcal{L}_{y,\lambda}(x) = \frac{1}{2}\|y - \Phi x\|_2^2 + \lambda \left\|D^* x\right\|_1.$$

We remind the reader that condition (H_0) is supposed to hold true in all our statements.

A.1. Preparatory lemmata

The following key lemma will be central in our proofs. It gives the first-order necessary and sufficient optimality conditions for the analysis variational problem $(P_\lambda(y))$.

(a) x_{ML} (b) $x^\star_\lambda(y)$ at the optimal λ (c) x_0



(d)

Fig. 2. Illustration of the selection of λ by minimizing GSURE_Π in a compressed sensing problem ($Q/N = 0.5$) by an ℓ^1-analysis regularization in a shift-invariant Haar wavelet dictionary. (a) The MLE x_{ML}. (b) A solution $x^\star_\lambda(y)$ of $(P_\lambda(y))$ at the optimal λ (the one minimizing GSURE_Π). (c) The underlying true image x_0. (d) Projection risk Risk_Π and its GSURE_Π estimate obtained from (10) using $k = 1$ random realization.

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

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

with

$$\Sigma_{y,\lambda}\big(x^\star_\lambda(y)\big) = \big\{\sigma \in \mathbb{R}^{|J|} \setminus \Phi^*\big(\Phi x^\star_\lambda(y) - y\big) + \lambda D_I s_I + \lambda D_J \sigma = 0 \text{ and } \|\sigma\|_\infty \leqslant 1\big\}, \tag{12}$$

where $I = J^c$ is the D-support of $x^\star_\lambda(y)$ and $s = \mathrm{sign}(D^* x^\star_\lambda(y))$.

Proof. The subdifferential of a real-valued proper convex function $F : \mathbb{R}^N \to \mathbb{R} \cup \{\infty\}$ is denoted ∂F. From standard convex analysis, we recall of ∂F at a point x in the domain of F

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

It is clear from this definition that $x^\star_\lambda(y)$ is a (global) minimizer of F if, and only if, $0 \in \partial F(x)$. By classical subdifferential calculus, the subdifferential of $\mathcal{L}_{y,\lambda}$ at x is the non-empty convex compact set

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

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

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

Taking $\sigma = u_J$, this is equivalent to $\sigma \in \Sigma_{y,\lambda}(x^\star_\lambda(y))$. □

The following lemma gives an implicit equation satisfied by any (non-necessarily unique) minimizer $x^\star_\lambda(y)$ of $(P_\lambda(y))$.

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

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

Proof. Owing to the first-order necessary and sufficient optimality condition (Lemma 1), there exists $\sigma \in \Sigma_{y,\lambda}(x_\lambda^\star(y))$ satisfying

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

By definition, $x_\lambda^\star(y) \in \mathcal{G}_J = (\operatorname{Im} D_J)^\perp$. We can then write $x_\lambda^\star(y) = U\alpha$ for some $\alpha \in \mathbb{R}^{\dim(\mathcal{G}_J)}$. Since $U^* D_J = 0$, multiplying both sides of (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, the implicit equation of $x_\lambda^\star(y)$ follows immediately. □

Suppose now that a vector satisfies the above implicit equation. The next lemma derives two equivalent necessary and sufficient conditions to guarantee that this vector is actually a solution to $(P_\lambda(y))$.

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

$$x_\lambda^\star(y) = \Gamma^{[J]}\Phi^* y - \lambda \Gamma^{[J]} D_I s_I,$$

where $s = \operatorname{sign}(D^* x_\lambda^\star(y))$. Then, $x_\lambda^\star(y)$ is a solution of $(P_\lambda(y))$ if, and only if, there exists $\sigma \in \mathbb{R}^{|J|}$ satisfying one of the following equivalent conditions

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

or

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

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. First, we observe that $x_\lambda^\star(y) \in \mathcal{G}_J$. According to Lemma 1, $x_\lambda^\star(y)$ is a solution of $(P_\lambda(y))$ if, and only if, there exists $\sigma \in \Sigma_{y,\lambda}(x_\lambda^\star(y))$. Since (H_J) holds, $\Gamma^{[J]}$ is properly defined. We can then plug the assumed implicit equation in (12) to get

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

Rearranging the terms multiplying y and s_I, we arrive at

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

This shows that $x_\lambda^\star(y)$ is a minimizer of $(P_\lambda(y))$ if, and only if

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

To prove the equivalence with (16), we first note that $U^*\tilde{\Omega}^{[J]} = 0$ implying that $\operatorname{Im}(\tilde{\Omega}^{[J]}) \subseteq \operatorname{Im}(D_J)$. Since $D_J D_J^+$ is the orthogonal projector on $\operatorname{Im}(D_J)$, we get $\tilde{\Omega}^{[J]} = D_J D_J^+\tilde{\Omega}^{[J]} = D_J\Omega^{[J]}$. With a similar argument, we get $\tilde{\Pi}^{[J]} = D_J\Pi^{[J]}$. Hence, the existence of $\sigma \in \Sigma_{y,\lambda}(x_\lambda^\star(y))$ such that $\|\sigma\|_\infty \leqslant 1$ is equivalent to

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

which in turn is equivalent to

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

We now show that even if $(P_\lambda(y))$ admits several solutions $x_\lambda^\star(y)$, all of them share the same image under Φ, which in turn implies that $y \mapsto \mu_\lambda^\star(y)$ is a single-valued mapping.

Lemma 4. *If x_1 and x_2 are two minimizers of $(P_\lambda(y))$, then $\Phi x_1 = \Phi x_2$.*

Proof. Let x_1, x_2 be two minimizers of $(P_\lambda(y))$. Suppose that $\Phi x_1 \neq \Phi x_2$. Take $x_3 = \rho x_1 + (1-\rho)x_2$, $\rho \in (0,1)$. Strict convexity of $u \mapsto \|y - u\|_2^2$ implies that

$$\frac{1}{2}\|y - \Phi x_3\|_2^2 < \frac{\rho}{2}\|y - \Phi x_1\|_2^2 + \frac{1-\rho}{2}\|y - \Phi x_2\|_2^2.$$

Jensen's inequality again applied to the ℓ^1-norm gives

$$\left\|D^*x_3\right\|_1 \leqslant \rho\left\|D^*x_1\right\|_1 + (1-\rho)\left\|D^*x_2\right\|_1.$$

Together, these two inequalities yield $\mathcal{L}_{y,\lambda}(x_3) < \mathcal{L}_{y,\lambda}(x_1)$, which contradicts our initial assumption that x_1 is a minimizer of $(P_\lambda(y))$. □

A.2. Proof of Theorem 1

Proof. Let $(y,\lambda) \notin \mathcal{H}$. By construction, the vector $x^\star_{\bar\lambda}(\bar y)$ obeys $D^*_J x^\star_{\bar\lambda}(\bar y) = 0$. Accordingly, for $(\bar y, \bar\lambda)$ sufficiently close to (y,λ), one has

$$\operatorname{sign}\big(D^*x^\star_{\bar\lambda}(\bar y)\big) = \operatorname{sign}\big(D^*x^\star_\lambda(y)\big).$$

Since $x^\star_\lambda$ is a solution of $(P_\lambda(y))$, using Lemmas 2 and 3, there exists σ such that

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

Let us split $J = K \cup L$, $K \cap L = \emptyset$ such that $\|\sigma_K\|_\infty = 1$ and $\|\sigma_L\|_\infty < 1$. Note that $\sigma_K \in \{-1,1\}^{|K|}$.

We first suppose that $\operatorname{Im}\tilde\Pi^{[J]} \subseteq \operatorname{Im} D_L$. To prove that $x^\star_{\bar\lambda}(\bar y)$ 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 take $\bar\sigma_L$ 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,

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

Since $\operatorname{Im}\tilde\Pi^{[J]} \subseteq \operatorname{Im} D_L$ and $D_LD_L^+$ is the orthogonal projector on $\operatorname{Im}(D_L)$, we have $\tilde\Pi^{[J]} = D_LD_L^+\tilde\Pi^{[J]}$. It follows that,

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

Now, for $(\bar y, \bar\lambda)$ close enough to (y,λ), we have

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

whence we deduce that $x^\star_{\bar\lambda}(\bar y)$ is a solution of $(\mathcal{P}_{\bar\lambda}(\bar y))$.

In fact, for $(y,\lambda) \notin \mathcal{H}$, we inevitably have $\operatorname{Im}\tilde\Pi^{[J]} \subseteq \operatorname{Im} D_L$. Indeed, projecting (17) on $\mathcal{G}_L$ gives

$$0 = \mathrm{P}_{\mathcal{G}_L}\big(\tilde\Pi^{[J]}y - \lambda\tilde\Omega^{[J]}s_I + \lambda D_J\sigma\big) = \mathrm{P}_{\mathcal{G}_L}\big(\tilde\Pi^{[J]}y - \lambda\tilde\Omega^{[J]}s_I + \lambda D_K\sigma_K\big),$$

or equivalently

$$\mathrm{P}_{\mathcal{G}_L}\tilde\Pi^{[J]}y = \mathrm{P}_{\mathcal{G}_L}\lambda\big(\tilde\Omega^{[J]}s_I - D_K\sigma_K\big).$$

If $\operatorname{Im}\tilde\Pi^{[J]} \not\subseteq \operatorname{Im} D_L$, then $(y,\lambda) \in \mathcal{H}_{J,K,s_I,\sigma_K}$, a contradiction. This concludes the proof. □

A.3. Proof of Theorem 2

Proof of (i). First it is easy to see that $\mathcal{H}_{J,K,s_{J^c},\sigma_K}$ in Definition 4 is a vector subspace of $\mathbb{R}^Q \times \mathbb{R}$. Moreover $\mathcal{H}_{J,K,s_{J^c},\sigma_K} \subseteq \ker \mathrm{P}_{\mathcal{G}_{J\setminus K}} B$, where $B = [\tilde{\Pi}^{[J]} \quad -\tilde{\Omega}^{[J]} s_{J^c} + D_K \sigma_K]$.

Now, fix λ. $\mathcal{H}_{\cdot,\lambda}$ is included in

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

where

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

Since $\operatorname{Im}\tilde{\Pi}^{[J]} \not\subset \operatorname{Im}(D_{J\setminus K})$, $\mathcal{H}^{\lambda}_{J,K,s_{J^c},\sigma_K}$ is an affine subspace of $\mathbb{R}^Q$ with $\dim(\tilde{\mathcal{H}}^{\lambda}_{J,K,s_{J^c},\sigma_K}) = \dim(\ker \mathrm{P}_{\mathcal{G}_{J\setminus K}} \tilde{\Pi}^{[J]}) < Q$, where the inequality follows from the rank-nullity theorem and the fact that $\mathcal{G}_{J\setminus K}$ is a (non-empty) strict subspace of $\mathbb{R}^Q$. Given that $\tilde{\mathcal{H}}^{\lambda}$ is a finite union of subspaces $\tilde{\mathcal{H}}^{\lambda}_{J,K,s_{J^c},\sigma_K}$ all strictly included in $\mathbb{R}^Q$, $\tilde{\mathcal{H}}^{\lambda}$ has a Lebesgue measure zero and so does $\mathcal{H}_{\cdot,\lambda} \subseteq \tilde{\mathcal{H}}^{\lambda}$.

Note that with a similar reasoning, one can show that $\mathcal{H}$ is also of zero Lebesgue measure using the fact that $B \notin \operatorname{Im} D_{J\setminus K}$ if $\tilde{\Pi}^{[J]} \notin \operatorname{Im} D_{J\setminus K}$ since $\operatorname{Im}\tilde{\Pi}^{[J]} \subseteq \operatorname{Im} B$. □

Proof of (ii). The proof of this statement is constructive. Denote by $\mathcal{M}_\lambda(y)$ the set of minimizers of $(P_\lambda(y))$. To lighten the notation, we drop the dependence on y and λ from $x^\star_\lambda(y) \in \mathcal{M}_\lambda(y)$.

First step. We prove the following statement

$$\left(\left(x^\star \in \mathcal{M}_\lambda(y) \wedge \neg(H_{\operatorname{supp}(D^*x^\star)^c})\right) \Longrightarrow \exists x^{\star\star}_\lambda(y) \in \mathcal{M}_\lambda(y) \wedge \operatorname{supp}\left(D^*x^{\star\star}\right) \subsetneq \operatorname{supp}\left(D^*x^\star\right)\right),$$

where $\wedge$ and $\neg$ are respectively the logical conjunction and negation symbols. In plain words, let $x^\star$ be a solution of $(P_\lambda(y))$. Suppose (H_J) does not hold where J is the D-cosupport of $x^\star$. We prove that there exists a solution $x^{\star\star}_\lambda(y)$ 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

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

$\mathcal{B}$ is a non-empty convex set and $0 \in \mathcal{B}$. Moreover for all $t \in \mathcal{B}$, $\partial\mathcal{L}_{y,\lambda}(v_t) = \partial\mathcal{L}_{y,\lambda}(x^\star)$. It then follows from Lemma 1 that for all $t \in \mathcal{B}$, v_t is a solution of $(P_\lambda(y))$. As a consequence, using Lemma 4, we get

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

Since $\lim_{|t|\to\infty} \|D^* v_t\|_1 = +\infty$, the set $\mathcal{B}$ is bounded. It is also an open set as a finite intersection of P open sets corresponding to the solutions to $\operatorname{sign}((D^*x^\star)_i + tz_i) = \operatorname{sign}((D^*x^\star)_i)$. Hence, $\mathcal{B}$ is an open interval of $\mathbb{R}$ which contains 0, i.e. there exist $t_1, t_0 \in \mathbb{R}$ such that

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

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

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

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

Second step. We now prove our claim, i.e.

$$\exists x^\star \in \mathcal{M}_\lambda(y) \text{ such that } (H_{\operatorname{supp}(D^*x^\star)^c}) \text{ holds.}$$

Consider $(x^\star_{(1)}, \dots, x^\star_{(P+1)}) \in (\mathcal{M}_\lambda(y))^{P+1}$ such that for every $i \in \{1, \dots, P+1\}$, the condition (H_{J_i}) does not hold for $J_i = \operatorname{supp}(D^*x^\star_{(i)})^c$ and $J_1 \supsetneq J_2 \supsetneq \cdots \supsetneq J_{P+1}$. Then, we have a strictly increasing sequence of $P+1$ subsets of $\{1, \dots, P\}$ which is impossible. Hence, according to the first step of our proof, there exists $i \in \{1, \dots, P+1\}$ such that (H_{J_i}) holds. □

Proof of (iii). By virtue of statement (ii), there exists a solution $x^\star_\lambda(y)$ of $(P_\lambda(y))$ such that (H_J) holds. Let us consider this solution. Using Theorem 1 for $\bar{y}$ close enough to y, we have

$$\Phi x_\lambda^\star(\bar{y}) = \Phi \Gamma^{[J]} \Phi^* \bar{y} - \lambda \Phi \Gamma^{[J]} D_I s_I,$$

where J is the D-cosupport of $x_\lambda^\star(y)$. Since I (hence J) and s_I are locally constant under the assumptions of the theorem, so is the vector $\lambda \Phi \Gamma^{[J]} D_I s_I$, it follows that $\mu_\lambda^\star(\bar{y}) = \Phi x_\lambda^\star(\bar{y})$ can be written as

$$\mu_\lambda^\star(\bar{y}) = \mu_\lambda^\star(y) + \Phi \Gamma^{[J]} \Phi^* (\bar{y} - y),$$

whence we deduce

$$\frac{\partial \mu_\lambda^\star(y)}{\partial y} = \Phi \Gamma^{[J]} \Phi^*.$$

Moreover, owing to statement (i), this expression is valid on $\mathbb{R}^Q \setminus \mathcal{H}_{\cdot,\lambda}$, a set of full Lebesgue measure. □

A.4. Proof of Theorem 3

We first recall Stein's lemma whose proof can be found in [27].

Lemma 5 *(Stein's lemma). Let $y = \Phi x_0 + w$ with $w \sim \mathcal{N}(0, \sigma^2 \mathrm{Id}_Q)$. Assume that $g : y \mapsto g(y)$ is weakly differentiable (and a fortiori a single-valued mapping), then*

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

Let us now turn to the proof of Theorem 3.

Proof of Theorem 3. Since $y \mapsto \hat{\mu}_\theta(y) = \Phi \hat{x}_\theta(y)$ is weakly differentiable, so is $A^* A \hat{\mu}_\theta(y)$ and we have

$$\frac{\partial A^* A \hat{\mu}_\theta(y)}{\partial y} = A^* A \frac{\partial \hat{\mu}_\theta(y)}{\partial y}.$$

Then, using Lemma 5, we get

$$\mathbb{E}_w \langle w, A^* A \hat{\mu}_\theta(y) \rangle = \sigma^2 \mathbb{E}_w \operatorname{tr}\left(A^* A \frac{\partial \hat{\mu}_\theta(y)}{\partial y} \right) = \sigma^2 \mathbb{E}_w \widehat{df}_\theta^A(y).$$

Using the decomposition $Ay = A\Phi x_0 + Aw$, we obtain

$$\begin{aligned}
\mathbb{E}_w \left\| Ay - A\hat{\mu}_\theta(y) \right\|_2^2 &= \mathbb{E}_w \|A\Phi x_0 + Aw\|_2^2 - 2\mathbb{E}_w \langle A\Phi x_0 + Aw, A\hat{\mu}_\theta(y) \rangle + \mathbb{E}_w \left\| A\hat{\mu}_\theta(y) \right\|_2^2 \\
&= \mathbb{E}_w \|A\Phi x_0\|_2^2 + \sigma^2 \operatorname{tr}(A^* A) - 2\mathbb{E}_w \langle A\Phi x_0, A\hat{\mu}_\theta(y) \rangle \\
&\quad - 2\mathbb{E}_w \langle w, A^* A \hat{\mu}_\theta(y) \rangle + \mathbb{E}_w \left\| A\hat{\mu}_\theta(y) \right\|_2^2 \\
&= \mathbb{E}_w \left\| A\Phi x_0 - A\hat{\mu}_\theta(y) \right\|_2^2 + \sigma^2 \operatorname{tr}(A^* A) - 2\sigma^2 \mathbb{E}_w \widehat{df}_\theta^A(y).
\end{aligned}$$

Moreover, $\sum_i \operatorname{cov}_w((Ay)_i, (A\hat{\mu}_\theta(y))_i) = \mathbb{E}_w \langle Aw, A\hat{\mu}_\theta(y) \rangle$, which shows that $\widehat{df}_\theta^A(y)$ is indeed an unbiased estimator of df_θ^A. □

A.5. Proof of Theorem 4

Proof. Denote by R^A the reliability of the GSURE for the estimator $\hat{x}_\theta(y)$, i.e.

$$R^A = \mathbb{E}_w \left(\mathrm{GSURE}^A(\hat{x}_\theta(y)) - \mathrm{SE}^A(\hat{x}_\theta(y)) \right)^2.$$

Let $Q^A(\hat{x}_\theta(y))$ be the quantity defined as

$$Q^A(\hat{x}_\theta(y)) = \|A\mu_0\|_2^2 + \left\| A\hat{\mu}_\theta(y) \right\|_2^2 - 2\langle Ay, A\hat{\mu}_\theta(y) \rangle + 2\sigma^2 \widehat{df}_\theta^A(y).$$

We have $\mathrm{GSURE}^A(\hat{x}_\theta(y)) - Q^A(\hat{x}_\theta(y)) = \|Ay\|_2^2 - \mathbb{E}_w \|Ay\|_2^2$, where

$$\mathbb{E}_w \|Ay\|_2^2 = \|A\mu_0\|_2^2 + \sigma^2 \operatorname{tr}(A^* A) \quad \text{and} \quad \mathbb{V}_w \|Ay\|_2^2 = 2\sigma^4 \left(\operatorname{tr}\left[(A^* A)^2 \right] + 2 \frac{\|A^* A \mu_0\|_2^2}{\sigma^2} \right).$$

It results that $\mathbb{E}_w(\mathrm{GSURE}^A(\hat{x}_\theta(y)) - Q^A(\hat{x}_\theta(y))) = 0$, and hence

$$\mathbb{E}_w\big(Q^A\big(\hat{x}_\theta(y)\big)\big) = \mathbb{E}_w\big(\mathrm{GSURE}^A\big(\hat{x}_\theta(y)\big)\big) = \mathbb{E}_w\big(\mathrm{SE}^A\big).$$

Remark that $Q^A(\hat{x}_\theta(y)) - \mathrm{SE}^A(\hat{x}_\theta(y)) = 2(\sigma^2 \widehat{df}_\theta^A(y) - \langle Aw, A\hat{\mu}_\theta(y)\rangle)$. We can now rewrite the reliability in the following form

$$\begin{aligned} R^A &= \mathbb{E}_w\big(\mathrm{GSURE}^A\big(\hat{x}_\theta(y)\big) - Q^A\big(\hat{x}_\theta(y)\big) + Q^A\big(\hat{x}_\theta(y)\big) - \mathrm{SE}^A\big(\hat{x}_\theta(y)\big)\big)^2 \\ &= \mathbb{V}_w\|Ay\|_2^2 + \mathbb{E}_w\big(Q^A\big(\hat{x}_\theta(y)\big) - \mathrm{SE}^A\big(\hat{x}_\theta(y)\big)\big)^2 + 4\underbrace{\mathbb{E}_w\big(\|Ay\|_2^2\big(\sigma^2\widehat{df}_\theta^A(y) - \big\langle Aw, A\hat{\mu}_\theta(y)\big\rangle\big)\big)}_{=T}. \end{aligned}$$

Lemma 5 gives $\mathbb{E}_w\langle Aw, A\hat{\mu}_\theta(y)\rangle = \sigma^2\mathbb{E}_w\widehat{df}_\theta^A(y)$, and we get

$$T = \underbrace{2\mathbb{E}_w\big(\langle Aw, A\mu_0\rangle\big(\sigma^2\widehat{df}_\theta^A(y) - \big\langle Aw, A\hat{\mu}_\theta(y)\big\rangle\big)\big)}_{T_1} + \underbrace{\mathbb{E}_w\big(\|Aw\|_2^2\big(\sigma^2\widehat{df}_\theta^A(y) - \big\langle Aw, A\hat{\mu}_\theta(y)\big\rangle\big)\big)}_{T_2}.$$

Let $\mu_A^\star(y) = A^*A\hat{\mu}_\theta(y)$, $\mu_A^0 = A^*A\mu_0$ and $w_A = A^*Aw$. Observe that $\widehat{df}_\theta^A(y) = \operatorname{div}\mu_A^\star(y)$ and $w_i(\mu_A^\star(y))_i$ is weakly differentiable. Then by integration by parts (in the same vein as in the proof of Stein's Lemma 5), we get

$$\begin{aligned} T_1 &= 2\sigma^2\sum_{i,j}\mathbb{E}_w\left(w_i\big(\mu_A^0\big)_i\frac{\partial\mu_A^\star(y)_j}{\partial w_j}\right) - 2\sum_{i,j}\mathbb{E}_w\big(w_i\big(\mu_A^0\big)_i w_j\mu_A^\star(y)_j\big) \\ &= -2\sigma^2\sum_{i,j}\mathbb{E}_w\left(\big(\mu_A^0\big)_i\frac{\partial w_i}{\partial w_j}\mu_A^\star(y)_j\right) = -2\sigma^2\mathbb{E}_w\big\langle\mu_A^0, \mu_A^\star(y)\big\rangle \end{aligned}$$

and

$$\begin{aligned} T_2 &= \sigma^2\sum_{i,j}\mathbb{E}_w\left(w_i(w_A)_i\frac{\partial\mu_A^\star(y)_j}{\partial w_j}\right) - \sum_{i,j}\mathbb{E}_w\big(w_i(w_A)_i w_j\mu_A^\star(y)_j\big) \\ &= -\sigma^2\sum_{i,j}\mathbb{E}_w\left(\frac{\partial w_i(w_A)_i}{\partial w_j}\mu_A^\star(y)_j\right) \\ &= -\sigma^2\sum_{j}\mathbb{E}_w\left(\mu_A^\star(y)_j\left(\sum_i\frac{\partial w_i(w_A)_i}{\partial w_j}\right)\right). \end{aligned}$$

In turn,

$$\sum_i\frac{\partial w_i(w_A)_i}{\partial w_j} = \frac{\partial}{\partial w_j}\|Aw\|_2^2 = 2\big(A^*Aw\big)_j = 2(w_A)_j.$$

Hence,

$$\begin{aligned} T_2 &= -2\sigma^2\sum_j\mathbb{E}_w\big(\big(\mu_A^\star(y)\big)_j(w_A)_j\big) = -2\sigma^2\mathbb{E}_w\big\langle w_A, \mu_A^\star(y)\big\rangle \\ &= -2\sigma^2\mathbb{E}_w\big\langle A^*Aw, A^*A\mu^\star(y)\big\rangle = -2\sigma^4\mathbb{E}_w\widehat{df}_\theta^{A^*A}(y), \end{aligned}$$

where the last equality is again a consequence of Lemma 5. It follows that $T = -2\sigma^2(\mathbb{E}_w\langle\mu_A^0, \mu_A^\star(y)\rangle + \sigma^2\mathbb{E}_w\widehat{df}_\theta^{A^*A}(y))$. Moreover using [56, Property 1] we have

$$\begin{aligned} \mathbb{E}_w\big(Q^A\big(\hat{x}_\theta(y)\big) - \mathrm{SE}^A\big(\hat{x}_\theta(y)\big)\big)^2 &= 4\mathbb{E}_w\big(\sigma^2\operatorname{div}\mu_A^\star(y) - \big\langle w, \mu_A^\star(y)\big\rangle\big)^2 \\ &= 4\sigma^2\left(\mathbb{E}_w\big\|\mu_A^\star(y)\big\|_2^2 + \sigma^2\mathbb{E}_w\operatorname{tr}\left[\left(\frac{\partial\mu_A^\star(y)}{\partial y}\right)^2\right]\right). \end{aligned}$$

Therefore, the reliability is given by

$$R^A = 2\sigma^4\operatorname{tr}\big[\big(A^*A\big)^2\big] + 4\sigma^2\mathbb{E}_w\big\|\mu_A^0 - \mu_A^\star(y)\big\|_2^2 + 4\sigma^4\mathbb{E}_w\left(\operatorname{tr}\left[\left(\frac{\partial\mu_A^\star(y)}{\partial y}\right)^2\right] - 2\widehat{df}_\theta^{A^*A}(y)\right).$$

Rearranging the last term above, we obtain the derived expression. □

A.6. Proof of Corollary 1

Proof. Let $\lambda \in \mathbb{R}_+^*$. From Theorem 2(iii), $y \mapsto \Phi x_\lambda^\star(y)$ is differentiable almost everywhere and we can invoke Theorem 3 to derive the GSURE expressions.

We also observe that $V = \Phi \Gamma^{[J]} \Phi^*$ is the orthogonal projector on $\operatorname{Im} V = \Phi(\mathcal{G}_J)$, so that $\operatorname{tr} V = \dim(\operatorname{Im} V) = \operatorname{rank}(\Phi \mathrm{P}_{\mathcal{G}_J})$. Since Φ is injective on $\mathcal{G}_J$ under (H_J), it follows that $\operatorname{tr} V = \dim(\mathcal{G}_J)$. Hence, using Theorem 3 with $A = \mathrm{Id}$, Theorem 2(ii) and (7), it follows that $\dim(\mathcal{G}_J)$ is an unbiased estimator of $df(\lambda)$. □

A.7. Proof of Corollary 2

Proof. As $V = \Phi \Gamma^{[J]} \Phi^*$ is the orthogonal projector on $\Phi(\mathcal{G}_J)$, we have

$$\operatorname{tr}\big[AVA^*A(2\,\mathrm{Id} - V)A^*\big] \geqslant 0.$$

Moreover A^*A is Hermitian, hence $\operatorname{tr}[(A^*A)^2] = \|A^*A\|_F^2$, we obtain (with the notation of reliability Section A.5) the following upper bound of the

$$R^A \leqslant 2\sigma^4 \big\|A^*A\big\|_F^2 + 4\sigma^2 \|A\|^4 \mathbb{E}\big\|\mu_0 - \mu_\lambda^\star(y)\big\|_2^2$$

where $\|.\|_F$ is the Frobenius norm and $\|.\|$ the matrix spectral norm. Then, for $A \in \mathbb{R}^{M \times Q}$, using classical inequalities, we get

$$\big\|A^*A\big\|_F^2 \leqslant \operatorname{rank}(A)\|A\|^4 = \min(M, Q)\|A\|^4 \leqslant Q\,\|A\|^4.$$

Since $x_\lambda^\star(y)$ is a solution of $(P_\lambda(y))$, we have

$$\frac{1}{2}\big\|y - \mu_\lambda^\star(y)\big\|_2^2 \leqslant \mathcal{L}_{y,\lambda}\big(x_\lambda^\star(y)\big) \leqslant \mathcal{L}_{y,\lambda}(0) = \frac{1}{2}\|y\|_2^2.$$

Thus, using Jensen's inequality, we get

$$\begin{aligned}\mathbb{E}\big\|\mu_0 - \mu_\lambda^\star(y)\big\|_2^2 &\leqslant 2\mathbb{E}\big(\|\mu_0 - y\|_2^2 + \big\|y - \mu_\lambda^\star(y)\big\|_2^2\big)\\ &\leqslant 2\mathbb{E}\big(\|w\|_2^2 + \|y\|_2^2\big) \leqslant 2\big(\|\mu_0\|_2^2 + 2Q\sigma^2\big).\end{aligned}$$

Altogether, this yields the following upper bound

$$\frac{R^A}{\sigma^4 Q^2} \leqslant \|A\|^4\left(\frac{18}{Q} + \frac{8\|\mu_0\|_2^2}{\sigma^2 Q^2}\right).$$

Since $\|\mu_0\|_2 < \infty$, this concludes the proof. □

A.8. Proof of Proposition 1

Proof. We have

$$\operatorname{tr}\big[A\Phi\Gamma^{[J]}\Phi^*A^*\big] = \operatorname{tr}\big[\Phi\Gamma^{[J]}\Phi^*A^*A\big].$$

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

$$\Gamma^{[J]}\Phi^*z = \operatorname*{arg\,min}_{h \in \mathcal{G}_J} \|\Phi h - z\|_2^2.$$

Writing the KKT conditions of this problem leads to (9), where $\tilde{\nu}$ are the Lagrange multipliers. □

References

[1] A. Kirsch, An Introduction to the Mathematical Theory of Inverse Problems, Appl. Math. Sci., vol. 120, Springer-Verlag, 2011.
[2] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1992) 259–268.
[3] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, T. Pock, An introduction to total variation for image analysis, in: Theoretical Foundations and Numerical Methods for Sparse Recovery, in: Radon Series Comp. Appl. Math., vol. 9, De Gruyter, ISBN 978-3110226140, 2010, pp. 263–340.
[4] M. Nikolova, Local strong homogeneity of a regularized estimator, SIAM J. Appl. Math. 61 (2000) 633–658.
[5] W. Ring, Structural properties of solutions to total variation regularization problems, M2AN Math. Model. Numer. Anal. 34 (2000) 799–810.

[6] V. Caselles, A. Chambolle, M. Novaga, The discontinuity set of solutions of the TV denoising problem and some extensions, Multiscale Model. Simul. 6 (2008) 879–894.
[7] R. Tibshirani, Regression shrinkage and selection via the Lasso, J. Roy. Statist. Soc. Ser. B 58 (1996) 267–288.
[8] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1999) 33–61.
[9] M. Elad, P. Milanfar, R. Rubinstein, Analysis versus synthesis in signal priors, Inverse Problems 23 (2007) 947–968.
[10] E. Candès, Y. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal. 31 (2010) 59–73.
[11] M. Grasmair, Linear convergence rates for Tikhonov regularization with positively homogeneous functionals, Inverse Problems 27 (2011) 075014.
[12] S. Nam, M. Davies, M. Elad, R. Gribonval, The cosparse analysis model and algorithms, arXiv:1106.4987v1, 2011.
[13] S. Vaiter, G. Peyré, C. Dossal, M.J. Fadili, Robust sparse analysis regularization, Trans. Inform. Theory (2012), in press.
[14] Y. Lu, M. Do, A theory for sampling signals from a union of subspaces, IEEE Trans. Signal Process. 56 (2008) 2334–2345.
[15] J. Bonnans, A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, 2000.
[16] B. Mordukhovich, Sensitivity analysis in nonsmooth optimization, in: D.A. Field, V. Komkov (Eds.), Theoretical Aspects of Industrial Design, in: SIAM Vol. Appl. Math., vol. 58, 1992, pp. 32–46.
[17] M. Osborne, B. Presnell, B. Turlach, On the LASSO and its dual, J. Comput. Graph. Statist. (2000) 319–337.
[18] M. Osborne, B. Presnell, B. Turlach, A new approach to variable selection in least squares problems, IMA J. Numer. Anal. 20 (2000) 389.
[19] D. Malioutov, M. Cetin, A. Willsky, Homotopy continuation for sparse signal representation, in: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), vol. 5, IEEE, 2005, pp. 733–736.
[20] D. Donoho, Y. Tsaig, Fast solution of ℓ^1-norm minimization problems when the solution may be sparse, IEEE Trans. Inform. Theory 54 (2008) 4789–4812.
[21] R. Tibshirani, J. Taylor, The solution path of the generalized lasso, Ann. Statist. 39 (2011) 1335–1371.
[22] B. Efron, How biased is the apparent error rate of a prediction rule? J. Amer. Statist. Assoc. 81 (1986) 461–470.
[23] C.L. Mallows, Some comments on C_p, Technometrics 15 (1973) 661–675.
[24] H. Akaike, Information theory and an extension of the maximum likelihood principle, in: Second International Symposium On Information Theory, vol. 1, Springer-Verlag, 1973, pp. 267–281.
[25] G. Schwarz, Estimating the dimension of a model, Ann. Statist. 6 (1978) 461–464.
[26] G. Golub, M. Heath, G. Wahba, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics (1979) 215–223.
[27] C. Stein, Estimation of the mean of a multivariate normal distribution, Ann. Statist. 9 (1981) 1135–1151.
[28] K.-C. Li, From Stein's unbiased risk estimates to the method of generalized cross validation, Ann. Statist. 13 (1985) 1352–1377.
[29] B. Efron, The estimation of prediction error: Covariance penalties and cross-validation, J. Amer. Statist. Assoc. 99 (2004) 619–632.
[30] D. Donoho, I. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.
[31] T. Cai, H. Zhou, A data-driven block thresholding approach to wavelet estimation, Ann. Statist. 37 (2009) 569–595.
[32] T. Blu, F. Luisier, The SURE-LET approach to image denoising, IEEE Trans. Image Process. 16 (2007) 2778–2786.
[33] I.M. Johnstone, Wavelet shrinkage for correlated data and inverse problems: Adaptivity results, Statist. Sinica 9 (1999) 51–83.
[34] D. Van de Ville, M. Kocher, SURE-based non-local means, IEEE Signal Process. Lett. 16 (2009) 973–976.
[35] V. Duval, J.-F. Aujol, Y. Gousseau, A bias-variance approach for the non-local means, SIAM J. Imaging Sci. 4 (2011) 760–788.
[36] C.-A. Deledalle, V. Duval, J. Salmon, Non-local methods with shape-adaptive patches (NLM-SAP), J. Math. Imaging Vision (2011) 1–18.
[37] J. Rice, Choice of smoothing parameter in deconvolution problems, in: Contemp. Math., vol. 59, 1986, pp. 137–151.
[38] C. Vonesch, S. Ramani, M. Unser, Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint, in: ICIP, IEEE, 2008, pp. 665–668.
[39] 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.
[40] Y.C. Eldar, Generalized SURE for exponential families: Applications to regularization, IEEE Trans. Signal Process. 57 (2009) 471–481.
[41] H. Zou, T. Hastie, R. Tibshirani, On the "degrees of freedom" of the lasso, Ann. Statist. 35 (2007) 2173–2192.
[42] M. Kachour, C. Dossal, M. Fadili, G. Peyré, C. Chesneau, The degrees of freedom of penalized l1 minimization, Statist. Sinica (2012), in press, http://dx.doi.org/10.5705/ss.2011.281.
[43] R.J. Tibshirani, J. Taylor, Degrees of freedom in lasso problems, Technical Report, arXiv:1111.0653, 2012.
[44] J.-C. Pesquet, A. Benazza-Benyahia, C. Chaux, A SURE approach for digital signal/image deconvolution problems, IEEE Trans. Signal Process. 57 (2009) 4616–4632.
[45] V. Solo, A SURE-fired way to choose smoothing parameters in ill-conditioned inverse problems, in: IEEE Int. Conf. Image Process. (ICIP), vol. 3, IEEE, 1996, pp. 89–92.
[46] J. Ye, On measuring and correcting the effects of data mining and model selection, J. Amer. Statist. Assoc. (1998) 120–131.
[47] X. Shen, J. Ye, Adaptive model selection, J. Amer. Statist. Assoc. 97 (2002) 210–221.
[48] 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.
[49] R. Giryes, M. Elad, Y. Eldar, The projected GSURE for automatic parameter tuning in iterative shrinkage methods, Appl. Comput. Harmon. Anal. 30 (2011) 407–422.
[50] C. Deledalle, S. Vaiter, G. Peyré, M.J. Fadili, C. Dossal, Proximal splitting derivatives for risk estimation, in: 2nd International Workshop on New Computational Methods for Inverse Problems (NCMIP), Paris.
[51] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vision 20 (2004) 89–97.
[52] A. Chambolle, T. Pock, A first-order primal–dual algorithm for convex problems with applications to imaging, J. Math. Imaging Vision 40 (2011) 120–145.
[53] H. Raguet, M.J. Fadili, G. Peyré, Generalized forward–backward splitting, Technical Report, 2011, Preprint Hal-00613637.
[54] N. Pustelnik, C. Chaux, J.-C. Pesquet, Parallel ProXimal Algorithm for image restoration using hybrid regularization, IEEE Trans. Image Process. 20 (2011) 2450–2462.
[55] P.L. Combettes, J.-C. Pesquet, Primal–dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum monotone operators, Set-Valued Var. Anal. 20 (2012) 307–330.
[56] F. Luisier, The SURE-LET approach to image denoising, PhD thesis, École Polytechnique Fédérale de Lausanne, 2010.