

# UNBIASED RISK ESTIMATION FOR SPARSE ANALYSIS REGULARIZATION

Charles-Alban Deledalle, Samuel Vaiter, Gabriel Peyré

Jalal Fadili

Charles Dossal

CEREMADE, CNRS-Paris-Dauphine

GREYC, CNRS-ENSICAEN IMB, CNRS-Bordeaux 1

## ABSTRACT

In this paper, we propose a rigorous derivation of the expression of the projected Generalized Stein Unbiased Risk Estimator (GSURE) for the estimation of the (projected) risk associated to regularized ill-posed linear inverse problems using sparsity-promoting  $\ell^1$  penalty. The projected GSURE is an unbiased estimator of the recovery risk on the vector projected on the orthogonal of the degradation operator kernel. Our framework can handle many well-known regularizations including sparse synthesis- (e.g. wavelet) and analysis-type priors (e.g. total variation). A distinctive novelty of this work is that, unlike previously proposed  $\ell^1$  risk estimators, we have a closed-form expression that can be implemented efficiently once the solution of the inverse problem is computed. To support our claims, numerical examples on ill-posed inverse problems with analysis and synthesis regularizations are reported where our GSURE estimates are used to tune the regularization parameter.

**Index Terms**— Sparsity, analysis regularization, inverse problems, risk estimator, GSURE.

## 1. INTRODUCTION

**Problem statement.** Many problems in image processing can be cast as recovering an image  $f_0 \in \mathbb{R}^N$  from the forward observation model

$$y = \Phi_0 f_0 + w, \quad (1)$$

where it is assumed throughout that  $w \sim \mathcal{N}(0, \sigma^2 \text{Id}_P)$ , and  $\Phi_0 : \mathbb{R}^N \rightarrow \mathbb{R}^P$  is a bounded linear operator which is typically ill-behaved since it models an acquisition process that entails loss of information so that  $P \leq N$ . 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. 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  $f_0$ . This prior information accounts for the smoothness of the solution and can range from the uniform smoothness assumption to more complex geometrical priors. To capture the complexity of image structures, we have witnessed a great deal of investigation where researchers spanning a wide range of viewpoints have advocated the use of sparsity in inverse problems.

**Sparse regularization.** Embarking from the degradation model (1), a popular approach to find sparse solutions to linear inverse problems, is to solve the following convex but non-smooth problem

$$x^* \in \underset{x \in \mathbb{R}^Q}{\operatorname{argmin}} \frac{1}{2} \|y - \Phi x\|^2 + \lambda \|D^* x\|_1, \quad (\mathcal{P}(y))$$

where  $\Phi = \Phi_0 \Psi$ ,  $\Psi : \mathbb{R}^Q \mapsto \mathbb{R}^N$  is a redundant synthesis dictionary, that maps coefficients  $x$  to images  $f = \Psi x$ ,  $D^*$  is an analysis linear operator and  $\lambda > 0$  is the regularization parameter.

This formulation encompasses several well-known problems in sparse recovery:

- *Synthesis-type regularization*, e.g. [1, 2]: in which case  $D^* = \text{Id}_Q$ , and we are seeking a sparse set of coefficients  $x^*$  and the solution signal or image is synthesized from these representation coefficients  $f^* = \Psi x^*$ .
- *Analysis-type regularization*, e.g. [3, 4]: which seeks a signal or image  $f$  whose coefficients  $D^* f$  are sparse, with  $\Psi = \text{Id}_N$ . Some problems are inherently of analysis-type such as the total variation regularization where  $D^* = \nabla : \mathbb{R}^N \mapsto \mathbb{R}^{2N}$  computes a discretized gradient of the image (a 2-D vector field) [5]. Another example of analysis regularization is the Fused Lasso [6], where  $D$  is the concatenation of a 1-D discrete derivative and the (weighted) identity.

$(\mathcal{P}(y))$  is versatile enough to allow hybridization of analysis and synthesis regularizations, e.g. to structure the synthesis coefficients in  $\Psi$  in a way that their gradient is sparse.

**Solving  $(\mathcal{P}(y))$ .** Although it is convex, solving  $(\mathcal{P}(y))$  is rather challenging given its non-smoothness. Moreover, the  $\ell^1$ -norm is composed with  $D^*$  and therefore the proximity

This work has been supported by the European Research Council (ERC project SIGMA-Vision) and the French National Research Agency (project NatImages).

operator is not computable in closed-form for general  $D$ . 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 originating from the non-smooth optimization realm. For instance, we here use the relaxed Arrow-Hurwicz algorithm as revitalized in [7]. This algorithm achieves full splitting where all operators are applied separately: proximity operators of  $\frac{1}{2}\|\cdot\|^2$  and  $\lambda\|\cdot\|_1$ ,  $\Phi$ ,  $D$  and their adjoints. See [4] for more details.

### 1.1. Previous work

**Analysis vs synthesis regularization.** It is known that in general, analysis and synthesis prior formulations can differ significantly, unless  $\Psi = \text{Id}_N$  and  $D$  is orthogonal or vice versa. Some insights can be found in [3] on the relation and distinction between analysis and synthesis-based  $\ell^1$ -sparsity regularizations. Moreover, there has been a recent wave of interest in the theoretical guarantees of analysis sparse recovery, see e.g. [4] and references therein.

**Degrees of freedom (DOF).** Degrees of freedom (DOF) is a familiar phrase in statistics. DOF is often used to quantify the complexity of a statistical estimation procedure. However, there is no exact correspondence between the DOF and the number of parameters in the model. The SURE theory [8] gives a rigorous definition of DOF for any estimation procedure including non-linear ones.

For synthesis  $\ell^1$  regularization, an unbiased estimate of the DOF was established in the overdetermined case in [9], and extended to the general setting in [10]. An unbiased estimate of the DOF for analysis  $\ell^1$  regularization is given in [4], see also [11].

**Risk estimation.** The DOF plays an important role in model validation and selection, and its unbiased estimates provide unbiased estimates of the risk on the prediction  $\Phi_0 f^*$  through the SURE [8]. These unbiased estimates of the risk can serve as a basis for automatic ways to choose the parameters of the reconstruction algorithm, e.g.  $\lambda$  in  $(\mathcal{P}(y))$ .

For instance, the SURE has been extensively used in sparse denoising and deconvolution, e.g. [12, 13]. Recently, a generalization of the SURE (GSURE) [14] has been developed for noise models within the multivariate canonical exponential family. This allows to estimate the risk on a projected version of  $f^*$ . Indeed, in the scenario where  $\Phi_0$  is rank-deficient or redundant, the GSURE can at best estimate the risk in  $\ker(\Phi_0)^\perp = \text{Im}(\Phi_0^*)$ . The projected GSURE has been applied to (non-iterative) wavelet-vaguelet deconvolution [15], and to sparse synthesis regularization [16, 17] with the iterative soft-thresholding algorithm.

### 1.2. Contributions

The main contribution of this paper is a rigorous derivation of the exact expression of the projected GSURE associated to  $(\mathcal{P}(y))$ . To this end, we capitalize on our recent work on local affine parameterization of solutions to  $(\mathcal{P}(y))$  [4]. Unlike previously proposed projected GSURE computations in [16, 17], we have a closed-form expression that can be implemented efficiently once the solution of the inverse problem is computed. Moreover, our framework elegantly covers in a unified way synthesis and analysis sparse regularizations.

## 2. SPARSE ANALYSIS SENSITIVITY

**Preliminaries.** Analysis regularization enforces that sparse solutions of  $(\mathcal{P}(y))$  belong to some co-space  $\mathcal{G}_J = \ker(D_J^*)$  where  $J$  is the co-support of a solution  $x^*$

$$J = I^c \quad \text{where} \quad I = \text{supp}(D^*x^*).$$

To be able to study the local sensitivity of the solution w.r.t.  $y$ , it is important that the forward operator  $\Phi = \Phi_0\Psi$  is well-behaved over the sparsity enforcing space  $\mathcal{G}_J$ . This is formalized through the assumption

$$\text{Ker } \Phi \cap \mathcal{G}_J = \{0\}. \quad (H_J)$$

When  $(H_J)$  holds we define the following mapping

$$A^{[J]} = U (U^* \Phi^* \Phi U)^{-1} U^*$$

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

**Local sensitivity.** For  $\lambda > 0$ , there exists a finite union of hyperplanes  $\mathcal{H}_\lambda \subset \mathbb{R}^Q$ , such that the following theorem holds.

**Theorem 1.** *Let  $y \notin \mathcal{H}_\lambda$  and let  $x^*$  be a solution of  $(\mathcal{P}(y))$ . Let  $I = \text{Supp}(D^*x^*)$  and  $s = \text{sign}(D^*x^*)$ . Suppose that  $(H_J)$  holds. Define*

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

*There exists an open neighborhood  $\mathcal{B}$  of  $y$  such that for every  $\bar{y} \in \mathcal{B}$ ,  $\hat{x}(\bar{y})$  is a solution of  $(\mathcal{P}(\bar{y}))$ .*

The exact definition of  $\mathcal{H}_\lambda$  is quite technical, and for obvious space limitation, we refer to [4] for details and a proof.

## 3. SURE RISK ESTIMATOR

**Projected GSURE.** In the following, we denote  $\Pi$  the orthogonal projector on  $\ker(\Phi_0)^\perp = \text{Im}(\Phi_0^*)$ , where it is assumed that  $\ker(\Phi_0) \neq \{0\}$ . This projector is  $\Pi = \Phi_0^*(\Phi_0\Phi_0^*)^+\Phi_0$ , where  $^+$  stands for the Moore-Penrose pseudo-inverse.

We first notice that by strict convexity of the quadratic part, even if  $(\mathcal{P}(y))$  admits several solutions  $x^*$  (that are obviously function of  $y$ ), all  $f^* = \Psi x^*$  share the same image under  $\Pi$ . We can thus define unambiguously  $\mu(y) = \Pi f^*$  as a (single-valued) mapping of the observation  $y$ .

Denoting  $\mu_0(y) = \Phi_0^*(\Phi_0\Phi_0^*)^+y$  the maximum likelihood estimator, the projected GSURE is defined following [15] as:

$$\text{GSURE}(y) = \|\mu_0(y) - \mu(y)\|^2 - \sigma^2 \text{tr}((\Phi_0\Phi_0^*)^+) + 2\sigma^2 \text{div}_{\Phi_0^*y}(\mu(y)),$$

where  $\text{div}_{\Phi_0^*y}(\mu(y))$  is the divergence of  $\mu(\cdot)$  w.r.t.  $\Phi_0^*y$ . Note that this estimator is introduced in [15] in the context of wavelet-vaguelet deconvolution, but it is easily seen to hold for arbitrary inverse problems and regularizations as long as the mapping  $\mu$  is weakly differentiable in the sense of [8]. Note also that up to a constant, which does not depends on  $\lambda$ , our definition of GSURE is equivalent to the one introduced in [14, 16].

Theorem 1 immediately implies our main result, that allows to compute the GSURE.

**Theorem 2.** *Except on a subset of  $\mathbb{R}^N$  of measure zero, the mapping  $y \mapsto \mu(y)$  is  $C^\infty$ , and*

$$\text{div}_{\Phi_0^*y}(\mu(y)) = \text{tr}(\Pi\Psi A^{[J]}\Psi^*).$$

A direct consequence of this result is that the above expression of the GSURE is an unbiased estimate of the risk on  $\text{Im}(\Phi^*)$ , i.e.

$$\mathbb{E}_w(\|\Pi f_0 - \mu(y)\|^2) = \mathbb{E}_w(\text{GSURE}(y)).$$

**Numerical computation of the GSURE.** The following proposition gives a way to compute efficiently the divergence term which boils down to solving a linear system.

**Proposition 1.** *One has*

$$\text{tr}(\Pi\Psi A^{[J]}\Psi^*) = \mathbb{E}_Z(\langle \Psi\nu(Z), \mu_0(Z) \rangle) \quad (2)$$

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

$$\begin{pmatrix} \Phi^*\Phi & D_J \\ D_J^* & 0 \end{pmatrix} \begin{pmatrix} \nu \\ \tilde{\nu} \end{pmatrix} = \begin{pmatrix} \Phi^*z \\ 0 \end{pmatrix}. \quad (3)$$

*Proof.* We have

$$\text{tr}(\Pi\Psi A^{[J]}\Psi^*) = \text{tr}(\Phi A^{[J]}\Psi^*\Phi_0^*(\Phi_0\Phi_0^*)^+).$$

Hence denoting  $\nu(z) = A^{[J]}\Phi^*z$ , and using the fact that for any matrix  $U$ ,  $\text{tr}(U) = \mathbb{E}_Z(\langle Z, UZ \rangle)$ , we arrive at (2).

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

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

The closed-form solution to this problem is given by (3).  $\square$

In practice, the empirical mean estimator is replaced for the expectation in (2), hence giving

$$\frac{1}{k} \sum_{i=1}^k \langle \Psi\nu(z_i), \mu_0(z_i) \rangle \xrightarrow{\text{WLLN}} \text{tr}(\Pi\Psi A^{[J]}\Psi^*),$$

for  $k$  realizations  $z_i$  of  $Z$ . The numerical computation of  $\nu(z_i)$  is achieved by solving the symmetric linear system (3) with a fast conjugate gradient solver.

#### 4. NUMERICAL ILLUSTRATIONS

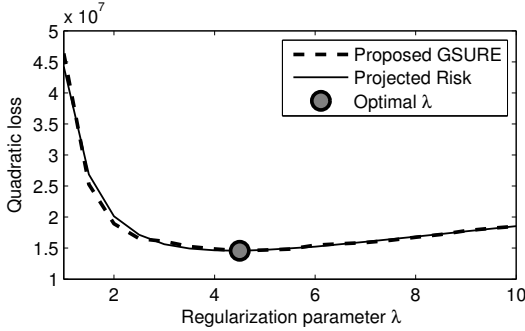
In this section, we exemplify the usefulness of our GSURE estimator which can serve as a basis for automatically tuning the value of  $\lambda$ . This is achieved by computing, from a single realization of the noise  $w$ , the parameter that minimizes the value of  $\text{GSURE}(y)$  for  $y = \Phi f_0 + w$ .

**Sparse Synthesis Regularization** In this example,  $\Phi_0$  is the (circulant) convolution operator associated to a Gaussian kernel of width 2 pixels, and the input PSNR is set to 27.78 dB. The regularization is of synthesis-type ( $D^* = \text{Id}_Q$ ) in an orthogonal wavelet dictionary  $\Psi$ . Fig. 1(a) depicts the projected risk and its GSURE estimate with  $k = 5$  as a function of  $\lambda$ . The curves are indeed unimodal and coincide even with  $k = 5$  and a single noise realization. Consequently, GSURE provides a high-quality estimate of  $\lambda$  minimizing the risk. A close in on the central area of the degraded and deconvolved (using the optimal  $\lambda$ ) images is shown in Fig. 1(b)-(c) for visual inspection of the restoration quality.

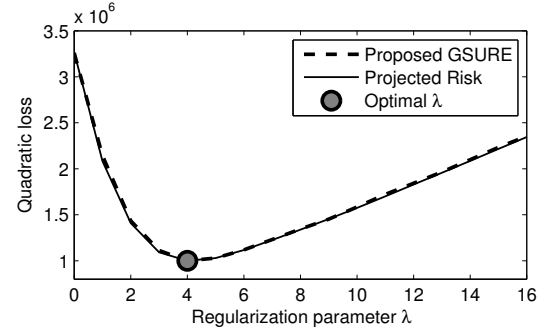
**Total Variation Regularization** We consider a compressed sensing setting where  $\Phi_0$  is a random partial DCT measurement matrix with an under-sampling ratio  $P/N = 0.5$ , and the PSNR is set to 27.50 dB. The regularization is of analysis-type, anisotropic total variation regularization ( $D^* = \nabla$ ,  $\Psi = \text{Id}_N$ ). The GSURE is estimated with  $k = 1$ . The results observed on the deconvolution example are confirmed in this compressed sensing experiment both visually and qualitatively, see Fig. 2.

#### 5. CONCLUSION

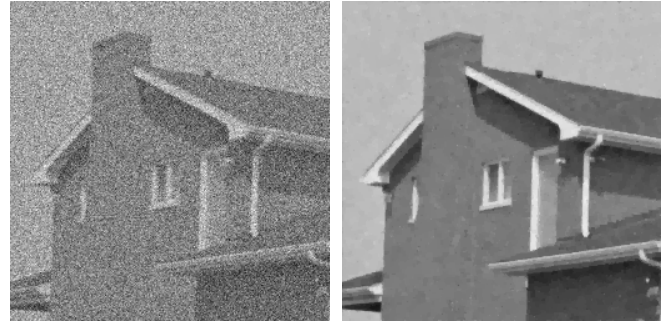
We proposed a grounded and computationally efficient framework to unbiasedly estimate the projected risk in  $\ell^1$ -regularized inverse problems handling both synthesis and analysis sparsity priors. Its usefulness has been illustrated on



(a)

(b)  $y$ (c)  $\mu(y)$  at the optimal  $\lambda$ 

(a)

(b)  $\mu_0(y)$ (c)  $\mu(y)$  at the optimal  $\lambda$ 

**Fig. 1.** Deconvolution problem (with a Gaussian PSF) with wavelet synthesis regularization, and  $\sigma = 10$ . (a) Projected risk and its GSURE estimate<sup>1</sup> using  $k = 5$  random realizations. (b)  $y$ . (c)  $\mu(y)$  at the optimal  $\lambda$ .

automatic choice of the regularization parameter for several inverse problem instances. The extension of this approach to other sparsity-promoting penalties, e.g. mixed norms for block structured sparsity, is currently under investigation.

## 6. REFERENCES

- [1] S.S. Chen, D.L. Donoho, and M.A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1999.
- [2] R. Tibshirani, “Regression shrinkage and selection via the Lasso,” *Journal of the Royal Statistical Society. Series B. Methodological*, vol. 58, no. 1, pp. 267–288, 1996.
- [3] M. Elad, P. Milanfar, and R. Rubinstein, “Analysis versus synthesis in signal priors,” *Inverse Problems*, vol. 23, no. 3, pp. 947–968, 2007.
- [4] S. Vaiter, G. Peyré, C. Dossal, and M.J. Fadili, “Robust sparse analysis regularization,” Tech. Rep., Preprint Hal-00627452, 2011.
- [5] L.I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [6] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, “Sparsity and smoothness via the fused Lasso,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.
- [7] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

<sup>1</sup>Without impacting the optimal choice of  $\lambda$ , the two curves have been vertically shifted to the same range of values for visualization purposes.

**Fig. 2.** Compressed sensing problem ( $P/N = 0.5$ ) with anisotropic total variation regularization ( $D^* = \nabla$ ,  $\Psi = \text{Id}_N$ ), and  $\sigma = 10$ . (a) Projected risk and its GSURE estimate<sup>1</sup> using  $k = 1$  random realization. (b)  $\mu_0(y)$ . (c)  $\mu(y)$  at the optimal  $\lambda$ .

- [8] C.M. Stein, “Estimation of the mean of a multivariate normal distribution,” *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [9] H. Zou, T. Hastie, and R. Tibshirani, “On the “degrees of freedom” of the lasso,” *The Annals of Statistics*, vol. 35, no. 5, pp. 2173–2192, 2007.
- [10] M. Kachour, C. Dossal, M.J. Fadili, G. Peyré, and C. Chesneau, “The degrees of freedom of penalized  $\ell_1$  minimization,” Tech. Rep., Preprint Hal-00638417, 2011, submitted.
- [11] R.J. Tibshirani and J. Taylor, “The solution path of the generalized Lasso,” *The Annals of Statistics*, vol. 39, no. 3, pp. 1335–1371, 2011.
- [12] F. Luisier, *The SURE-LET approach to image denoising*, Ph.D. thesis, École polytechnique fédérale de lausanne, 2010.
- [13] F.-X. Dupé, M.J. Fadili, and J.-L. Starck, “A proximal iteration for deconvolving Poisson noisy images using sparse representations,” *IEEE Transactions on Image Processing*, vol. 18, no. 2, 2009, 310–321.
- [14] Y. C. Eldar, “Generalized SURE for exponential families: Applications to regularization,” *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 471–481, 2009.
- [15] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, “A SURE approach for digital signal/image deconvolution problems,” *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4616–4632, 2009.
- [16] C. Vonesch, S. Ramani, and M. Unser, “Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint,” in *IEEE ICIP*. IEEE, 2008, pp. 665–668.
- [17] R. Giryes, M. Elad, and Y.C. Eldar, “The projected GSURE for automatic parameter tuning in iterative shrinkage methods,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–422, 2011.