



Block based refitting in ℓ_{12} sparse regularisation

Charles-Alban Deledalle, Nicolas Papadakis, Joseph Salmon, Samuel Vaiter

► To cite this version:

Charles-Alban Deledalle, Nicolas Papadakis, Joseph Salmon, Samuel Vaiter. Block based refitting in ℓ_{12} sparse regularisation. Journal of Mathematical Imaging and Vision, 2021, 63, pp.216-236. 10.1007/s10851-020-00993-2 . hal-02330441

HAL Id: hal-02330441

<https://hal.science/hal-02330441>

Submitted on 24 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Block based refitting in ℓ_{12} sparse regularisation

Charles-Alban Deledalle

Department of Electrical and Computer Engineering

University of California

San Diego, La Jolla, USA

Nicolas Papadakis

CNRS, Univ. Bordeaux, IMB, UMR 5251,

F-33400 Talence, France

Joseph Salmon

IMAG, Univ. Montpellier, CNRS, Montpellier, France

Samuel Vaiter

CNRS & IMB, UMR 5584

Univ. Bourgogne Franche-Comté, F-21000 Dijon, France

Abstract

In many linear regression problems, including ill-posed inverse problems in image restoration, the data exhibit some sparse structures that can be used to regularize the inversion. To this end, a classical path is to use ℓ_{12} block based regularization. While efficient at retrieving the inherent sparsity patterns of the data – the support – the estimated solutions are known to suffer from a systematical bias. We propose a general framework for removing this artifact by refitting the solution towards the data while preserving key features of its structure such as the support. This is done through the use of refitting block penalties that only act on the support of the estimated solution. Based on an analysis of related works in the literature, we introduce a new penalty that is well suited for refitting purposes. We also present a new algorithm to obtain the refitted solution along with the original (biased) solution for any convex refitting block penalty. Experiments illustrate the good behavior of the proposed block penalty for refitting solutions of Total Variation and Total Generalized Variation models.

1 Introduction

We consider linear inverse problems of the form $y = \Phi x + w$, where $y \in \mathbb{R}^p$ is an observed degraded image, $x \in \mathbb{R}^n$ the unknown clean image, $\Phi : \mathbb{R}^n \rightarrow \mathbb{R}^p$ a linear operator and $w \in \mathbb{R}^p$ a noise component, typically a zero-mean white Gaussian random vector with standard deviation $\sigma > 0$. To reduce the effect of noise and the potential ill-conditioning of Φ , we consider a regularized least squares problem with a sparse analysis regularization term based on an ℓ_{12} block penalty of the form

$$\hat{x} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \|\Gamma x\|_{1,2} \quad , \quad (1)$$

where $\lambda > 0$ is a regularization parameter, $\Gamma : \mathbb{R}^n \rightarrow \mathbb{R}^{m \times b}$ is a linear analysis operator mapping an image over m blocks of size b , and for $z \in \mathbb{R}^{m \times b}$

$$\|z\|_{1,2} = \sum_{i=1}^m \|z_i\|_2 = \sum_{i=1}^m \left(\sum_{j=1}^b z_{i,j}^2 \right)^{1/2}, \quad (2)$$

with $z_i = (z_{i,j})_{j=1}^b \in \mathbb{R}^b$. Note that the terminology *sparse analysis* is used here by opposition to *sparse synthesis* models as discussed in the seminal work of [21]. The first term in (1) is a data fidelity term enforcing x to be close to y through Φ , while the second term enforces the so-called group sparsity on x (sometimes referred to as joint sparsity, block sparsity or structured sparsity) capturing the organization of the data as encoded by Γ , see for instance [1].

1.1 Related examples

A typical example is the Lasso (Least absolute shrinkage and selection operator) [40]. The Lasso is a statistical procedure used for variable selection and relying on regularized linear least square regression as expressed in eq. (1) in which $\Gamma = \text{Id}$, $m = n$ and $b = 1$. The Lasso is known to promote sparse solutions, *i.e.*, such that $\hat{x}_k = 0$ for most indices $1 \leq k \leq n$. Since blocks are of size $b = 1$, the regularization term boils down to the classical ℓ_1 sparsity term that is unstructured as no interactions between the elements of \hat{x} are considered. In this paper, we will focus instead on cases of block penalties where $b > 1$. The group Lasso [48, 25] is one of them, for which Γ is designed to reorganize the elements \hat{x}_k into groups $\hat{z}_i = (\Gamma \hat{x})_i$ supposedly meaningful according to some prior knowledge on the data. The group Lasso is known to promote block sparse solutions, *i.e.*, such that $\hat{z}_i = 0_b$ for most of the groups $1 \leq i \leq m$. Note that elements within a non-zero group $\hat{z}_i \neq 0_b$ are not required to be sparse.

Regarding image restoration applications, the authors of [30] use an ℓ_{12} regularization term where Γ extracts $m = n$ overlapping blocks of wavelet coefficients of size b or where Γ use a dyadic decomposition of the wavelet coefficients into blocks of variable size but non-overlapping [31]. Such strategy was also used in audio processing [47] for denoising. Another example that we will investigate here is the one of the total-variation (TV) [35]. We can distinguish two different forms of TV models. Anisotropic total-variation (TV_{aniso}) [22], considers Γ the operator which concatenates the vertical and horizontal components of the discrete gradients into a vector of size $m = 2n$, hence $b = 1$. Isotropic total-variation (TV_{iso}) considers instead $\Gamma = \nabla$ being the operator which extracts $m = n$ discrete image gradient vectors of size $b = 2$. Unlike TV_{aniso}, TV_{iso} jointly enforces vertical and horizontal components of the gradient to be simultaneously zero. Since TV_{aniso} does not take into account interactions between both directions, it over favors vertical and horizontal structures while TV_{iso} behaves similarly in all directions, hence their name [22]. Both models promote sparsity of the discrete gradient field of the image, and, as a result, their solutions are piece-wise constant. A major difference is that TV_{aniso} favors constant regions that are rectangular-like shaped and separated by sharp edges, while TV_{iso} favors constant regions that are rounded-like shaped and separated by fast but gradual transitions [16].

As TV promotes piece-wise constant solutions, it is known to produce *staircasing* artifacts that are all the more harmful as the images contain shaded objects [18, 10]. To reduce this effect, the authors of [26] suggested using an ℓ_{12} block sparsity term not only by grouping vertical and horizontal components of the gradient, but by grouping neighboring gradients in overlapping patches. An alternative to reduce staircasing, that we will also investigate here, is the second order Total Generalized Variation (TGV) model [4] that promotes piece-wise affine solutions. As we will see, TGV is another example of models that falls into this type of least squares problems regularized with a sparse analysis term based on ℓ_{12} block penalties. Sparsity in that case encodes that sought images are composed of few shaded regions with few variations of slopes and separated by edges.

1.2 Support of the solution

Solutions of the sparse analysis regularization model in (1) are known to be sparse [21, 28], *i.e.*, such that $(\Gamma x)_i = 0_b$ for most blocks $1 \leq i \leq m$. It results that a key notion, central to all of these estimators, is the one of support, *i.e.*, the set of non-zero blocks in $\Gamma \hat{x}$, defined as

$$\hat{\mathcal{I}} = \text{supp}(\Gamma \hat{x}) = \{1 \leq i \leq m : (\Gamma \hat{x})_i \neq 0_b\} . \quad (3)$$

For the group Lasso, the support is typically used to identify groups of covariates (columns of Φ) being explanatory variables for the dependent variable y (*i.e.*, significantly correlated with y). For TViso, the support is the set of pixel indices where transitions occur in the restored image. In general, the support plays an important role as it captures the intrinsic structural information underlying the data. While being biased, in practice, the estimate \hat{x} obtained by sparse analysis regularization (1) recover quite correctly the support $\text{supp}(\Gamma x)$ of the underlying sparse signal x . Under some additional assumptions, support recovery is even proven to be exact as proved in [44] for $b = 1$ (anisotropic case) and [45] for $b \geq 1$.

1.3 Bias of the solution

Though the support $\hat{\mathcal{I}}$ of \hat{x} can be very close to the one of the sought image x , the estimated amplitudes \hat{x}_i suffers from a systematical bias. When $\Phi = \text{Id}$, the Lasso corresponds to the soft-thresholding (ST) operator

$$\hat{x}_i = \text{ST}(y_i, \lambda) = \max(y_i - \lambda \text{sign } y_i, 0) \quad (4)$$

for which all non-zero elements of the solutions are shrinked towards 0 by a shift $\pm\lambda$ resulting to under- and over-estimated values. With TViso, this bias is reflected by a loss of contrast in the image since the amplitudes of some regions are regressed towards the mean of the image [37, 43, 42]. In TGV, not only a loss of contrast results from this bias, but one can observe that the slopes in areas of transitions are often over estimated.

1.4 Boosting approaches

Given the artifacts induced by the ℓ_{12} sparse regularization, many approaches have been developed to re-enhance the quality of the solutions, *e.g.*, to reduce the loss of contrast and staircasing of TViso. We refer to these approaches as boosting. Most of them consist in solving (1) iteratively based on the residue $\Phi \hat{x} - y$, or a related quantity, obtained during the previous iterations. Among them, the well-known Bregman iterations [29] is often considered to recover part of the loss of contrast for TViso. Other related procedures are twicing [41], boosting with the ℓ_2 loss [6], unsharp residual iteration [11], SAIF-boosting [27, 39], ideal spectral filtering in the analysis sense [23] and SOS-boosting [34]. While these approaches reduce the bias in the estimated amplitudes, the support $\hat{\mathcal{I}}$ of the original solution is not guaranteed to be preserved in the boosted solution, even though this one may correspond to the support of the sought image x .

1.5 Projection on the support

Given the key role of the support of solutions of (1), we believe that it is of main importance that a re-enhanced solution \tilde{x} preserves it, *i.e.*, such that $\text{supp}(\Gamma x) \subseteq \hat{\mathcal{I}}$. For this reason, we focus on re-fitting strategies that, unlike boosting, reduce the bias while preserving the support of the original solution. In the Lasso ($\Gamma = \text{Id}$, $m = n$ and $b = 1$) [40], a well known re-fitting scheme consists in performing *a posteriori* a least-square re-estimation of the non-zero coefficients of the solution. This post re-fitting technique became popular under various names in the statistical literature: Hybrid Lasso [20], Lasso-Gauss [33], OLS post-Lasso [3], Debiased Lasso (see [24, 3])

for extensive details on the subject). Such approaches consists in approximating y through Φ by an image sharing the same support as \hat{x} :

$$\hat{x}^{\text{supp}} \in \underset{x; \text{supp}(\Gamma x) \subseteq \hat{\mathcal{I}}}{\text{argmin}} \quad \frac{1}{2} \|\Phi x - y\|_2^2, \quad (5)$$

where $\hat{\mathcal{I}} = \text{supp}(\Gamma \hat{x})$. While this strategy works well for blocks of size $b = 1$, *e.g.*, for the Lasso or TVaniso, it suffers from an excessive increase of variance whenever $b \geq 2$ (see [16] for illustrations on TViso). This is due to the fact that solutions do not only present sharp edges, but may involve gradual transitions. To cope with this issue, additional features of \hat{x} than its support must be also preserved by a refitting procedure.

1.6 Advanced refitting strategies

For the Lasso, it has been observed that a pointwise preservation of the sign of \hat{x}_i onto the support improves the numerical performances of the refitting [12]. For $b = 2$ and TViso like models, the joint projection on the support with conservation of the direction (or orientation) of $(\Gamma \hat{x})_i$ has been proposed in [5]. Extension to second order regularization such as TGV [4] are investigated in [8] in the context of partially order spaces and approximate operators Φ . In a parallel line of research, it has been proposed in [46] to respect the inclusion of the level lines of \hat{x} in the refitting by solving an isotonic regression problem. All these models are constrained to respect exactly the orientation $(\Gamma \hat{x})_i$ of the biased solution on elements of the support, *i.e.*, when $i \in \hat{\mathcal{I}}$. In [16, 32], an alternative approach, based on the preservation of covariant information between \hat{x} and y , aims only at preserving the orientation $(\Gamma \hat{x})_i$ to some extent. While also respecting the support of \hat{x} , this gives more flexibility for the refitted solution to correct \hat{x} and adapt to the data content y . This model is nevertheless insensitive to the direction and it involves a quadratic penalty that tends to promote over smoothed refittings.

1.7 Outline and contributions

In Section 2, we present a general framework for refitting solutions promoted by ℓ_{12} sparse regularization (1) that extends a preliminary version of this work [17]. Our variational refitting method relies on the use of block penalty functions that act on the support of the biased solution \hat{x} . We introduce the Soft-penalized Direction model (SD), while discussing suitable properties a refitting block penalty should satisfy.

In Section 3, we propose stable algorithms to compute our refitting strategy for any convex refitting block penalty.

We show in Section 4 how our model relates and inherits the advantages of other methods such as Bregman iterations [29] or de-biasing approaches [5, 16].

Experiments in Section 5 exhibit the practical benefits for the SD refitting for imaging problems involving TViso, a variant of TViso for color images and TGV based regularization.

2 Refitting with block penalties

The refitting procedure of a biased solution \hat{x} of (1) is expressed in the following general framework

$$\tilde{x}^\phi \in \underset{x; \text{supp}(\Gamma x) \subseteq \hat{\mathcal{I}}}{\text{argmin}} \quad \frac{1}{2} \|\Phi x - y\|_2^2 + \sum_{i \in \hat{\mathcal{I}}} \phi((\Gamma x)_i, (\Gamma \hat{x})_i), \quad (6)$$

where $\phi : \mathbb{R}^b \times \mathbb{R}^b \rightarrow \mathbb{R}$ is a block penalty ($b \geq 1$ is the size of the blocks) promoting Γx to share information with $\Gamma \hat{x}$ in some sense to be specified. To compute global optimum of the refitting model (6), we only consider in this paper refitting block penalties such that $z \mapsto \phi(z, \hat{z})$ is convex.

To refer to some features of the vector $\Gamma\hat{x}$, let us first define properly the notions of relative orientation, direction and projection between two vectors.

Definition 1. Let z and \hat{z} be vectors in \mathbb{R}^b , we define

$$\cos(z, \hat{z}) = \left\langle \frac{z}{\|z\|_2}, \frac{\hat{z}}{\|\hat{z}\|_2} \right\rangle = \frac{1}{\|z\|_2 \|\hat{z}\|_2} \sum_{j=1}^b z_j \hat{z}_j, \quad (7)$$

$$\text{and } P_{\hat{z}}(z) = \left\langle z, \frac{\hat{z}}{\|\hat{z}\|_2} \right\rangle \frac{\hat{z}}{\|\hat{z}\|_2} = \frac{\|z\|_2}{\|\hat{z}\|_2} \cos(z, \hat{z}) \hat{z}, \quad (8)$$

where $P_{\hat{z}}(z)$ is the orthogonal projection of z onto $\text{Span}(\hat{z})$ (i.e., the orientation axis of \hat{z}). We say that z and \hat{z} share the same orientation (resp. direction), if $|\cos(z, \hat{z})| = 1$ (resp. $\cos(z, \hat{z}) = 1$). We also consider that $\cos(z, \hat{z}) = 1$ in case of null vectors $z = 0_b$ and/or $\hat{z} = 0_b$.

Thanks to Definition 1, we can now introduce our refitting block penalty designed to preserve the desired features of $\hat{z} = \Gamma\hat{x}$ in a simple way. We call our block penalty the *Soft-penalized Direction* (SD) penalty which reads as

$$\phi_{\text{SD}}(z, \hat{z}) = \lambda \|z\|_2 (1 - \cos(z, \hat{z})) . \quad (9)$$

We also introduce five other alternatives, the Hard constrained Orientation (HO) penalty

$$\phi_{\text{HO}}(z, \hat{z}) = \iota_{\{z \in \mathbb{R}^b: |\cos(z, \hat{z})|=1\}}(z), \quad (10)$$

where $\iota_{\mathcal{C}}$ is the $0/+ \infty$ indicator function of a set \mathcal{C} , the Hard-constrained Direction (HD) penalty

$$\phi_{\text{HD}}(z, \hat{z}) = \iota_{\{z \in \mathbb{R}^b: \cos(z, \hat{z})=1\}}(z), \quad (11)$$

the Quadratic penalized Orientation (QO) penalty

$$\phi_{\text{QO}}(z, \hat{z}) = \frac{\lambda \|z\|_2^2}{2 \|\hat{z}\|_2^2} (1 - \cos^2(z, \hat{z})) , \quad (12)$$

the Quadratic penalized Direction (QD) penalty

$$\phi_{\text{QD}}(z, \hat{z}) = \begin{cases} \frac{\lambda}{2} \frac{\|z\|_2^2}{\|\hat{z}\|_2^2} (1 - \cos^2(z, \hat{z})) & \text{if } \cos(z, \hat{z}) \geq 0 \\ \frac{\lambda}{2} \frac{\|z\|_2^2}{\|\hat{z}\|_2^2} & \text{otherwise} \end{cases} \quad (13)$$

and the Soft-constrained Orientation (SO) penalty

$$\phi_{\text{SO}}(z, \hat{z}) = \lambda \|z\|_2 \sqrt{1 - \cos^2(z, \hat{z})} . \quad (14)$$

We will see in Section 4 that HD, HO and QO lead us to retrieve existing refitting models known respectively in the litterature as ICB (Infimal Convolution between Bregman distances) debiasing [5], Bregman debiasing [5], and CLEAR (Covariant LEAst square Refitting) [16].

2.1 Desired properties of refitting block penalties

We now introduce properties a block penalty ϕ should satisfy for refitting purposes, for any \hat{z} :

- (P1) ϕ is convex, non negative and $\phi(z, \hat{z}) = 0$, if $\cos(z, \hat{z}) = 1$ or $\|z\|_2 = 0$,
- (P2) $\phi(z', \hat{z}) \geq \phi(z'', \hat{z})$ if $\|z'\|_2 = \|z''\|_2$ and $\cos(z'', \hat{z}) \geq \cos(z', \hat{z})$,
- (P3) $z \mapsto \phi(z, \hat{z})$ is continuous,

Table 1: Properties satisfied by the considered block penalties.

Properties	HO	HD	QO	QD	SO	SD
1	✓	✓	✓	✓	✓	✓
2		✓		✓		✓
3			✓	✓	✓	✓
4					✓	✓
a	✓	✓	✓	✓	✓	✓
b	✓	✓	✓	✓	✓	✓
c		✓		✓		✓
d		✓		✓		✓
e			✓	✓	✓	✓

(P4) $\phi(z, \hat{z}) \leq C\|z\|_2$, for $C > 0$.

Property (P1) stipulates that no configuration can be more favorable than z and \hat{z} having the same direction. Hence, the direction of the refitted solution should be encouraged to follow the one of the biased solution. Property (P2) imposes that for a fixed amplitude, the penalty should be increasing w.r.t. the angle formed with \hat{z} . Property (P3) enforces refitting that can continuously adapt to the data and be robust to small perturbations. Property (P4) claims that a refitting block penalty should not penalize more some configurations than the original penalty $\|\cdot\|_{1,2}$, at least up to some multiplicative factor $C > 0$.

The next proposition, whose proof is given in Appendix A, provides additional properties.

Proposition 1. *Properties of block penalties lead to the following implications.*

- (a) $(P1) \Rightarrow \|\Phi\tilde{x}^\phi - y\|_2^2 \leq \|\Phi\hat{x} - y\|_2^2$.
- (b) $(P1) \Rightarrow \phi$ is non decreasing with respect to $\|z\|_2$ for a fixed angle (z, \hat{z}) ,
- (c) $(P2) \Rightarrow \phi$ is symmetric with respect to the orientation axis induced by \hat{z} ,
- (d) $(P1) + (P2) \Rightarrow \phi(z', \hat{z}) \geq \phi(z, \hat{z})$, if $\|z'\|_2 \geq \|z\|_2$ and $\cos(z, \hat{z}) = \cos(z', \hat{z})$,
- (e) $(P1) + (P3) \Rightarrow \phi(z, \hat{z}) \rightarrow 0$ when $\cos(z, \hat{z}) \rightarrow 1$,

2.2 Properties of considered block penalties

The properties of the previously introduced refitting block penalties are synthesized in Table 1. The proposed SD model is the only one satisfying all the desired properties. As illustrated in Fig. 1, it is a continuous penalization that increases continuously with respect to the absolute angle between z and \hat{z} .

Other block penalties are insensitive to directions (HO, QO and SO), completely intolerant (HD and HO) or too tolerant (QD) to small changes of orientations, hence not satisfying. These drawbacks will be illustrated in our experiments conducted in Section 5.

When $b = 1$, the orientation-based penalties (QO, HO and SO) have absolutely no effect while the direction-based penalties HD and QD preserve the sign of $(\Gamma\hat{x})_i$. In this paper, when $b \geq 1$, we argue that the direction of the block $(\Gamma\hat{x})_i$ carries important information that is worth preserving when refitting, at least to some extent.

3 Refitting in practice

We now introduce a general algorithm aiming to jointly solve the original problem (1) and the refitting one (6) for any refitting block penalty ϕ . This framework has been extended from the

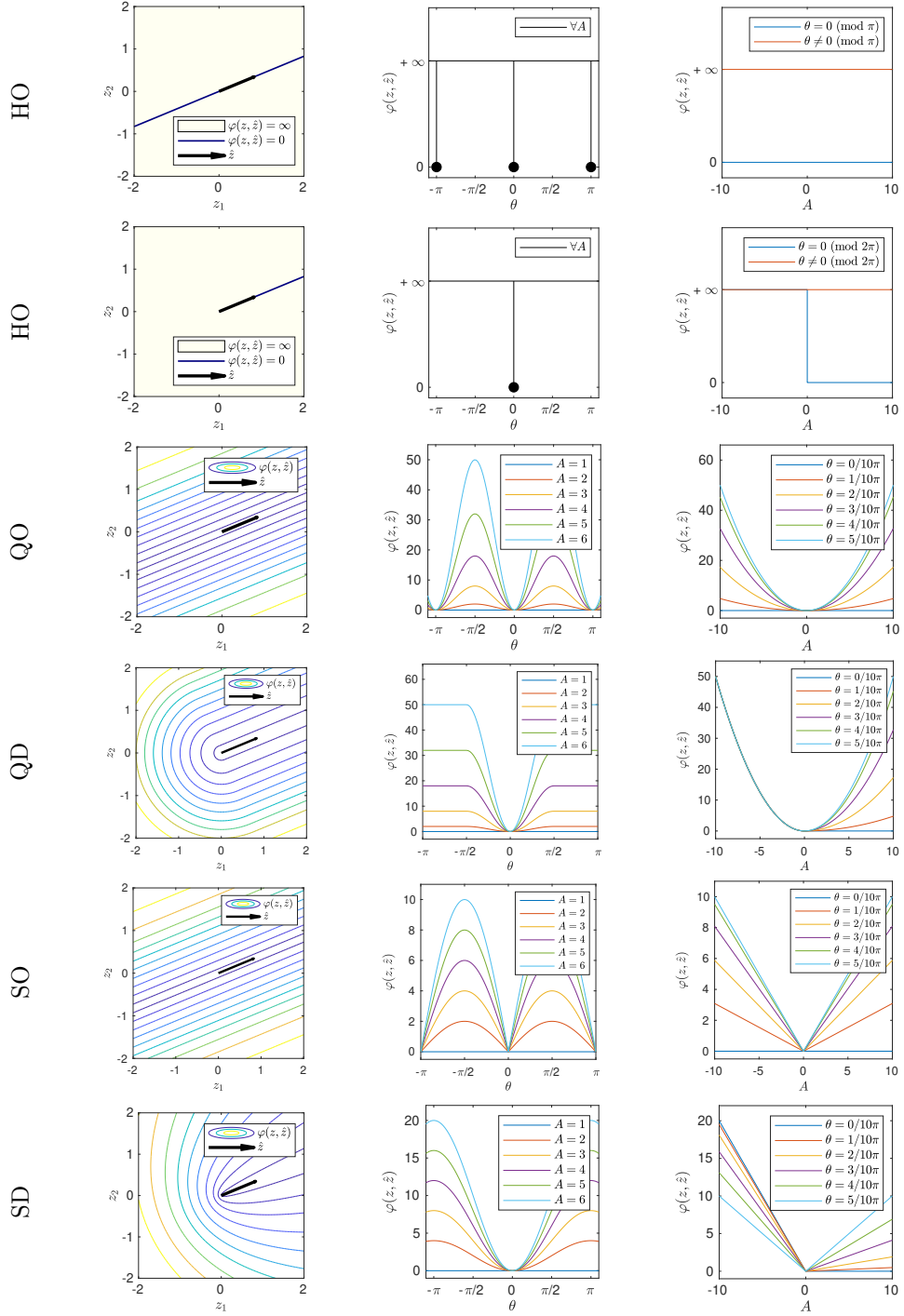


Figure 1: Illustration of continuous block penalties QO and SD: (left) 2D level lines of ϕ for $z = (z_1, z_2) = A \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \hat{z}$, (middle) evolution regarding θ and (right) A .

stable projection onto the support developed in [15] and later adapted to refitting with the

Quadratic Orientation penalty in [16].

Given \hat{x} solution of (1), a posterior refitting can be obtained by solving (6) for any refitting block penalty ϕ . To that end, we write the characteristic function of support preservation as $\sum_{i \in \hat{\mathcal{I}}^c} \iota_{\{0\}}(\xi_i)$, where $\iota_{\{0\}}(\xi) = 0$ if $z = 0$ and $+\infty$ otherwise. By introducing the convex function

$$\omega_\phi(\xi, \Gamma \hat{x}, \hat{\mathcal{I}}) = \sum_{i \in \hat{\mathcal{I}}^c} \iota_{\{0\}}(\xi_i) + \sum_{i \in \hat{\mathcal{I}}} \phi(\xi_i, (\Gamma \hat{x})_i) , \quad (15)$$

the general refitting problem (6) can be expressed as

$$\tilde{x}^\phi \in \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|\Phi x - y\|_2^2 + \omega_\phi(\Gamma x, \Gamma \hat{x}, \hat{\mathcal{I}}) . \quad (16)$$

We now describe two iterative algorithms that can be used for the joint computation of \hat{x} and \tilde{x} .

3.1 Primal-dual formulation

We first consider the primal dual formulation of the problems (1) and (16) that reads

$$\min_{x \in \mathbb{R}^n} \max_{\xi \in \mathbb{R}^b} \frac{1}{2} \|\Phi x - y\|_2^2 + \langle \Gamma x, \xi \rangle - \iota_{B_\lambda^2}(\xi) , \quad (17)$$

$$\min_{x \in \mathbb{R}^n} \max_{\xi \in \mathbb{R}^b} \frac{1}{2} \|\Phi x - y\|_2^2 + \langle \Gamma x, \xi \rangle - \omega_\phi^*(\xi, \Gamma \hat{x}, \hat{\mathcal{I}}) , \quad (18)$$

where $\iota_{B_\lambda^2}$ is the indicator function of the ℓ_2 ball of radius λ (that is 0 if $\|z_i\|_2 \leq \lambda$ for all $i \in [m]$ and $+\infty$ otherwise) and

$$\omega_\phi^*(\xi, \Gamma \hat{x}, \hat{\mathcal{I}}) = \sup_{\zeta \in \mathbb{R}^b} \langle \xi, \zeta \rangle - \omega_\phi(\zeta, \Gamma \hat{x}, \hat{\mathcal{I}}) \quad (19)$$

is the convex conjugate, with respect to the first argument of $\omega_\phi(\cdot, \Gamma \hat{x}, \hat{\mathcal{I}})$.

Two iterative primal-dual algorithms are used to solve these problems. They involve the biased variables (\hat{z}^k, \hat{x}^k) and the refitted ones $(\tilde{z}^k, \tilde{x}^k)$. Let us now present the whole algorithm, defined for parameters $\kappa > 0$, $\tau > 0$ and $\theta \in [0, 1]$ as:

$$\begin{cases} \hat{\nu}^{k+1} &= \hat{z}^k + \kappa \Gamma \hat{v}^k \\ \tilde{\nu}^{k+1} &= \tilde{z}^k + \kappa \Gamma \tilde{v}^k \\ \hat{\xi}^{k+1} &= \Pi(\hat{\nu}^{k+1}, \lambda) \\ \hat{\mathcal{I}}^{k+1} &= \{i \in [m] : \|\hat{\nu}_i^{k+1}\|_2 > \lambda + \beta\} \\ \tilde{\xi}^{k+1} &= \operatorname{prox}_{\kappa \omega_\phi^*}(\tilde{\nu}^{k+1}, \Psi(\hat{\nu}^{k+1}), \hat{\mathcal{I}}^{k+1}) \\ \hat{x}^{k+1} &= \Phi_\tau^- \left(\hat{x}^k + \tau(\Phi^t y - \Gamma^t \hat{\xi}^{k+1}) \right) \\ \tilde{x}^{k+1} &= \Phi_\tau^- \left(\tilde{x}^k + \tau(\Phi^t y - \Gamma^t \tilde{\xi}^{k+1}) \right) \\ \hat{v}^{k+1} &= \hat{x}^{k+1} + \theta(\hat{x}^{k+1} - \hat{x}^k) \\ \tilde{v}^{k+1} &= \tilde{x}^{k+1} + \theta(\tilde{x}^{k+1} - \tilde{x}^k), \end{cases} \quad (20)$$

with the operator $\Phi_\tau^- = (\operatorname{Id} + \tau \Phi^t \Phi)^{-1}$ and Π , $\operatorname{prox}_{\kappa \omega_\phi^*}$ and Ψ are functions that will be detailed bellow. As we will see in Proposition 2, the quantity $\beta > 0$ is used to guarantee that we estimate the support of \hat{x} correctly. In practice, we choose β as the smallest available non-zero floating number. The process for the biased variable involves the orthogonal projection on the ℓ_2 ball of radius λ in \mathbb{R}^b

$$\Pi(\hat{\xi}, \lambda)_i = \frac{\lambda \hat{\xi}_i}{\max(\lambda, \|\hat{\xi}_i\|_2)} .$$

whereas the refitting algorithm relies on the proximal operator of ω_ϕ^* . We recall that the proximal operator of a convex function ψ at point ξ_0 reads

$$\text{prox}_{\kappa\psi}(\xi) = \underset{\xi}{\text{argmin}} \frac{1}{2\kappa} \|\xi - \xi_0\|_2^2 + \psi(\xi). \quad (21)$$

From the block structure of the function ω_ϕ defined in (15), the computation of its proximal operator may be realized pointwise. Since $\iota_{\{0\}}(\xi)^* = 0$, we have

$$\text{prox}_{\kappa\omega_\phi^*}(\xi^0, \Gamma\hat{x}, \hat{\mathcal{I}})_i = \begin{cases} \text{prox}_{\kappa\phi^*}(\xi_i^0, (\Gamma\hat{x})_i), & \text{if } i \in \hat{\mathcal{I}}, \\ \xi_i^0, & \text{otherwise.} \end{cases} \quad (22)$$

Table 2 gives the expressions of the dual functions ϕ^* with respect to their first variable and their related proximal operators $\text{prox}_{\kappa\phi^*}$ for the refitting block penalties considered in this paper. All details are given in the Appendix B.

Following [9], for any positive scalars τ and κ satisfying $\tau\kappa\|\Gamma^t\Gamma\|_2 < 1$ and $\theta \in [0, 1]$, the estimates $(\hat{\xi}^k, \hat{x}^k, \hat{v}^k)$ of the biased solution converge to $(\hat{\xi}, \hat{x}, \hat{v})$, where $(\hat{\xi}, \hat{x}, \hat{v})$ is a saddle point of (17). When the last arguments of the function ω_ϕ^* are the converged $\Gamma\hat{x}$ and its support $\hat{\mathcal{I}}$, the refitted variables converge to a saddle point of (18). However, we did not succeed to show convergence for the refitting process, since the quantities $\Gamma\hat{x}$ are $\hat{\mathcal{I}}$ are only estimated from the biased variables at the current iteration, as explained in the next paragraphs.

Online support identification. Estimating $\text{supp}(\Gamma\hat{x})$ from an estimation \hat{x}^k is not stable numerically: the support $\text{supp}(\Gamma\hat{x}^k)$ can be far from $\text{supp}(\Gamma\hat{x})$ even though \hat{x}^k is arbitrarily close to \hat{x} . As in [15], we rather consider the dual variable $\hat{\xi}^k$ to estimate the support. From relations in (35), we indeed expect at convergence $\hat{\xi}^k$ to saturate on the support of $\Gamma\hat{x}$ and to satisfy the optimality condition $\hat{\xi}_i^k = \lambda \frac{(\Gamma\hat{x})_i}{\|(\Gamma\hat{x})_i\|_2}$. In practice, the norm of the dual variable $\hat{\xi}_i^k$ saturates to λ relatively fast onto $\hat{\mathcal{I}}$. As a consequence, it is far more stable to detect the support of $\Gamma\hat{x}$ with the dual variable $\hat{\xi}^k$ than with the vector $\Gamma\hat{x}^k$ itself. The next proposition, adapted from [15] and whose sketch of proof is given in Appendix A, shows that this approach can indeed converge towards the support $\hat{\mathcal{I}}$.

Proposition 2. *Let $\alpha > 0$ be the minimum non zero value of $\|(\Gamma\hat{x})_i\|_2$, and choose β such that $\alpha\kappa > \beta > 0$. Then, denoting $\hat{v}^{k+1} = \hat{\xi}^k + \kappa\Gamma\hat{v}^k$, $\hat{\mathcal{I}}^{k+1} = \{i \in [m] : \|\hat{v}_i^{k+1}\|_2 > \lambda + \beta\}$ in Algorithm (20) converges in finite time to the true support $\hat{\mathcal{I}} = \text{supp}(\Gamma\hat{x})$ of the biased solution¹.*

Online direction and norm identification. In algorithm (20), the function $\Psi(\hat{v}^{k+1})$ aims at approximating $\Gamma\hat{x}$. As for the estimation of the support, instead of directly considering $\Gamma\hat{x}^k$, we rather rely on $\hat{v}^{k+1} = \hat{\xi}^k + \kappa\Gamma\hat{v}^k$ to obtain a stable estimation \hat{z} of the vector $\Gamma\hat{x}$.

Table 2 shows that for all the considered block penalties, the computation of the proximal operator of ϕ^* involves the normalized vector $\hat{z}_i/\|\hat{z}_i\|_2$ on the support $\hat{\mathcal{I}}$. This direction is approximated at each iteration by normalizing \hat{v}_i^{k+1} . The amplitude $\|\hat{z}_i\|_2$ is just required for the models QO and QD.

In Proposition 7, we will show that the model of [16] is related to the QO block penalty. This method considers the algorithmic differentiation of the biased process to obtain a refitting algorithm. Given its good numerical performance, we leverage this approach to define the function Ψ as shown in the next proposition whose proof can be found in Appendix A.

Proposition 3. *The approach of [16] corresponds to refit with the QO block penalty and the following convergent estimation $\Psi(\hat{v}_i^{k+1})$ of $(\Gamma\hat{x})_i$.*

$$\Psi(\hat{v}_i^{k+1}) = \frac{\|\hat{v}_i^{k+1}\|_2 - \lambda}{\kappa\|\hat{v}_i^{k+1}\|_2} \hat{v}_i^{k+1}. \quad (23)$$

¹As in [5], the extended support $\|\hat{\xi}_i\|_2 = \lambda$ can be tackled by testing $\|\hat{v}_i^{k+1}\|_2 \geq \lambda$.

Discussion. This joint-estimation considers at every iteration k different refitting functions $\omega_\phi^*(\cdot, \Psi(\hat{\nu}^{k+1}), \hat{\mathcal{I}}^{k+1})$ in (16). Then, unless $b = 1$ (see [15]), we cannot show the convergence of the refitting scheme. As in [16], we nevertheless observe convergence and a stable behavior for this algorithm.

In addition to its better numerical stability, the running time of joint-refitting is more interesting than the posterior approach. In Algorithm (20), the refitted variables at iteration k require the biased variables at the same iteration and the whole process can be realized in parallel without significantly affecting the running time of the original biased process. On the other hand, posterior refitting is necessarily sequential and the running time is doubled in general.

3.2 Douglas-Rachford formulation

An alternative to obtain solutions \hat{x} of (1) and \tilde{x} of (16) is to consider the splitting TViso reformulation, as proposed in [13], and given by

$$\min_{\substack{x \in \mathbb{R}^n \\ \xi \in \mathbb{R}^b}} \frac{1}{2} \|\Phi x - y\|_2^2 + \iota_{\{x, \xi; \Gamma x = \xi\}}(x, \xi) + \lambda \|\xi\|_{1,2} , \quad (24)$$

$$\min_{\substack{x \in \mathbb{R}^n \\ \xi \in \mathbb{R}^b}} \frac{1}{2} \|\Phi x - y\|_2^2 + \iota_{\{x, \xi; \Gamma x = \xi\}}(x, \xi) + \omega_\phi(\xi, \Gamma \hat{x}, \hat{\mathcal{I}}) . \quad (25)$$

This problem can be solved with the Douglas-Rachford algorithm [19]. Introducing the parameters $\alpha \in (0, 2)$ and $\tau > 0$, the iterates read

$$\begin{cases} \hat{v}^{k+1} = \Gamma^-(2\hat{x}^k - \hat{\mu}^k + \Gamma^t(2\hat{\xi}^k - \hat{\zeta}^k)), \\ \tilde{v}^{k+1} = \Gamma^-(2\tilde{x}^k - \tilde{\mu}^k + \Gamma^t(2\tilde{\xi}^k - \tilde{\zeta}^k)), \\ \hat{\mu}^{k+1} = \hat{\mu}^k + \alpha(\hat{v}^{k+1} - \hat{x}^k), \\ \tilde{\mu}^{k+1} = \tilde{\mu}^k + \alpha(\tilde{v}^{k+1} - \tilde{x}^k), \\ \hat{\zeta}^{k+1} = \hat{\zeta}^k + \alpha(\Gamma \hat{v}^{k+1} - \hat{\xi}^k), \\ \tilde{\zeta}^{k+1} = \tilde{\zeta}^k + \alpha(\Gamma \tilde{v}^{k+1} - \tilde{\xi}^k), \\ \hat{x}^{k+1} = \Phi_\tau^-(\hat{\mu}^{k+1} + \tau \Phi^t y), \\ \tilde{x}^{k+1} = \Phi_\tau^-(\tilde{\mu}^{k+1} + \tau \Phi^t y), \\ \hat{\xi}^{k+1} = \text{ST}(\hat{\zeta}^{k+1}, \tau \lambda), \\ \hat{\mathcal{I}}^{k+1} = \left\{ i \in [m] : \|\hat{\zeta}_i^{k+1}\|_2 > \tau \lambda + \beta \right\}, \\ \tilde{\xi}^{k+1} = \text{prox}_{\tau \omega_\phi(\cdot, \Upsilon(\hat{\zeta}^{k+1}), \hat{\mathcal{I}}^{k+1})}(\tilde{\zeta}^{k+1}) \end{cases} \quad (26)$$

with $\Gamma^- = (\text{Id} + \Gamma^t \Gamma)^{-1}$, the block Soft Thresholding (ST) operator

$$\text{ST}(\hat{\zeta}, \lambda)_i = \begin{cases} 0 & \text{if } \|\hat{\zeta}_i\|_2 \leq \lambda, \\ \hat{\zeta}_i - \lambda \frac{\hat{\zeta}_i}{\|\hat{\zeta}_i\|_2} & \text{otherwise} . \end{cases}$$

and the proximal operator of ω_ϕ that can be written pointwise as

$$\begin{aligned} & \text{prox}_{\tau \omega_\phi}(\zeta^0, \Gamma \hat{x}, \hat{\mathcal{I}})_i \\ &= \begin{cases} \zeta^0 - \tau \text{prox}_{\phi^*/\tau}(\zeta_i^0/\tau, (\Gamma \hat{x})_i), & \text{if } i \in \hat{\mathcal{I}} , \\ 0, & \text{otherwise} . \end{cases} \end{aligned} \quad (27)$$

The estimates $(\hat{x}^k, \hat{\xi}^k, \hat{\zeta}^k)$ of the biased solution converge to $(\hat{x}, \hat{\xi}, \hat{\zeta}) = (\hat{x}, \Gamma \hat{x}, (1 + \tau \lambda)(\Gamma \hat{x}))$, from the optimality conditions of problem (24). On the other hand, there is again no convergence guarantee for the refitted variables since the last two arguments of the function ω_ϕ are potentially modified at each iteration.

Table 2: Convex conjugates and proximal operators of the studied block penalties ϕ .

ϕ	$\phi^*(z, \hat{z})$	$\text{prox}_{\kappa\phi^*}(z_0, \hat{z})$
HO	$\begin{cases} 0, & \text{if } \cos(z, \hat{z}) = 0 \\ +\infty, & \text{otherwise} \end{cases}$	$z_0 - P_{\hat{z}}(z_0)$
HD	$\begin{cases} 0, & \text{if } \cos(z, \hat{z}) \leq 0 \\ +\infty, & \text{otherwise} \end{cases}$	$\begin{cases} z_0 - P_{\hat{z}}(z_0), & \text{if } \langle z_0, \hat{z} \rangle \geq 0 \\ z_0, & \text{otherwise} \end{cases}$
QO	$\begin{cases} \frac{\ \hat{z}\ _2}{2\lambda} \ z\ _2^2, & \text{if } \cos(z, \hat{z}) = 0 \\ +\infty, & \text{otherwise} \end{cases}$	$\frac{\lambda}{\lambda + \kappa \ \hat{z}\ _2} (z_0 - P_{\hat{z}}(z_0))$
QD	$\begin{cases} \frac{\ \hat{z}\ _2}{2\lambda} \ z\ _2^2 & \text{if } \cos \theta(z, \hat{z}) \leq 0 \\ +\infty & \text{otherwise} \end{cases}$	$\frac{\lambda}{\lambda + \kappa \ \hat{z}\ _2} \begin{cases} z_0 - P_{\hat{z}}(z_0) & \text{if } \langle z_0, \hat{z} \rangle \geq 0 \\ z_0 & \text{otherwise} \end{cases}$
SO	$\begin{cases} 0, & \text{if } \cos(z, \hat{z}) = 0 \text{ and } \ z\ _2 \leq \lambda \\ +\infty, & \text{otherwise} \end{cases}$	$\lambda \frac{z_0 - P_{\hat{z}}(z_0)}{\max(\lambda, \ z_0 - P_{\hat{z}}(z_0)\ _2)}$
SD	$\begin{cases} 0, & \text{if } \ z + \lambda \frac{\hat{z}}{\ \hat{z}\ _2}\ \leq \lambda^\dagger \\ +\infty, & \text{otherwise} \end{cases}$	$\lambda \left(\frac{z_0 + \lambda \frac{\hat{z}}{\ \hat{z}\ _2}}{\max(\lambda, \ z_0 + \lambda \frac{\hat{z}}{\ \hat{z}\ _2}\ _2)} - \frac{\hat{z}}{\ \hat{z}\ _2} \right)$

† : note that the condition implies that $\cos(z, \hat{z}) \leq 0$.

The support of $\Gamma\hat{x}$ is estimated in line from auxiliary variables as $\hat{\mathcal{I}}^{k+1} = \{i \in [m] : \|\hat{\zeta}_i^{k+1}\|_2 > \tau\lambda + \beta\}$, where, as shown in [14], β has to be taken such that $0 < \beta < \lambda \min_{i \in \hat{\mathcal{I}}} \|(\Gamma\hat{x})_i\|_2$ to have convergence in finite time of $\hat{\mathcal{I}}^{k+1}$ to the true support $\hat{\mathcal{I}} = \text{supp}(\Gamma\hat{x})$. Again the quantity $\beta > 0$ is chosen in practice as being the smallest available non-zero floating number.

Considering the stable algorithmic differentiation strategy of [16] suggested in the previous subsection, the vector $\Gamma\hat{x}$ is approximated on the support $\hat{\mathcal{I}}^{k+1}$ at each iteration with the function $\Upsilon(\hat{\zeta}) = \frac{\|\hat{\zeta}\|_2 - \lambda\tau}{\|\hat{\zeta}\|_2} \hat{\zeta}$.

4 Related re-fitting works

We now review some other related refitting methods. First we will discuss of refitting methods based on Bregman divergences, next the refitting approach developped in [16], and then we will see how these techniques are related to the block penalties introduced in Section 2.

4.1 Bregman-based Refitting

4.1.1 Bregman divergence of ℓ_{12} structured regularizers

In the literature [29, 5], Bregman divergences have proven to be well suited to measure the discrepancy between the biased solution \hat{x} and its refitting \tilde{x} . We recall that for a convex, proper and lower semicontinuous function ψ , the associated (generalized) Bregman divergence between x and \hat{x} is, for any subgradient $\hat{p} \in \partial\psi(\hat{x})$:

$$D_{\psi}^{\hat{p}}(x, \hat{x}) = \psi(x) - \psi(\hat{x}) - \langle \hat{p}, x - \hat{x} \rangle \geq 0. \quad (28)$$

If ψ is an absolutely 1-homogeneous function, i.e., $\psi(\alpha x) = |\alpha|\psi(x)$, $\forall \alpha \in \mathbb{R}$, then

$$p \in \partial\psi(x) \Rightarrow \psi(x) = \langle p, x \rangle, \quad (29)$$

and the Bregman divergence simplifies into

$$D_{\psi}^{\hat{p}}(x, \hat{x}) = \psi(x) - \langle \hat{p}, x \rangle. \quad (30)$$

As an example, let us consider $\psi(x) = \|x\|_2$. Since ψ is 1-homogeneous, it follows that

$$D_{\|\cdot\|_2}^{\hat{p}}(x, \hat{x}) = \|x\|_2 - \langle \hat{p}, x \rangle \quad (31)$$

$$\text{where } \hat{p} \in \partial\|\cdot\|_2(\hat{x}) \Leftrightarrow \begin{cases} \hat{p} = \frac{\hat{x}}{\|\hat{x}\|_2} & \text{if } \hat{x} \neq 0, \\ \|\hat{p}\|_2 \leq 1 & \text{otherwise.} \end{cases} \quad (32)$$

For regularizers of the form $\psi(x) = \|\Gamma x\|_{1,2}$, we introduce the following notations

$$\Omega(\hat{x}) = \partial\|\Gamma \cdot\|_{1,2}(\hat{x}), \quad (33)$$

$$\Delta(\hat{x}) = \{\hat{\eta} : \Gamma^t \hat{\eta} \in \partial\|\Gamma \cdot\|_{1,2}(\hat{x})\}. \quad (34)$$

We have, see for instance [7], for all $\hat{p} \in \mathbb{R}^n$:

$$\hat{p} \in \Omega(\hat{x}) \Leftrightarrow \exists \hat{\eta} \in \Delta(\hat{x}) \text{ s.t. } \hat{p} = \Gamma^t \hat{\eta} \quad (35)$$

or in short $\Omega(\hat{x}) = \Gamma^t \Delta(\hat{x})$. Interestingly $\Delta(x)$ enjoys a separability property in terms of all subgradients associated to the ℓ_2 norms of the blocks

$$\Delta(\hat{x}) = \{\hat{\eta} \in \mathbb{R}^m : \hat{\eta}_i \in \partial\|\cdot\|_2((\Gamma \hat{x})_i), \forall i \in [m]\}. \quad (36)$$

Additionally, the next proposition shows that there is a similar separability property for the Bregman divergence (refer to Appendix A for the proof).

Proposition 4. *Let $\hat{\eta} \in \Delta(\hat{x})$ and $\hat{p} = \Gamma^t \hat{\eta}$. Then*

$$D_{\|\Gamma \cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) = \sum_{i=1}^m D_{\|\cdot\|_2}^{\hat{\eta}_i}((\Gamma x)_i, (\Gamma \hat{x})_i). \quad (37)$$

In the following we consider \hat{x} and its support $\hat{\mathcal{I}}$ to be fixed, and we denote by $D_i^{\hat{\eta}}(\Gamma x)$ the following

$$D_i^{\hat{\eta}}(\Gamma x) = D_{\|\cdot\|_2}^{\hat{\eta}_i}((\Gamma x)_i, (\Gamma \hat{x})_i). \quad (38)$$

The next proposition shows that such a divergence measures the fit of directions between $(\Gamma x)_i$ and $(\Gamma \hat{x})_i$, but also partially captures the support $\hat{\mathcal{I}}$ (refer to Appendix A for the proof).

Proposition 5. *Let $\hat{\eta} \in \Delta(\hat{x})$. We have for $i \in \hat{\mathcal{I}}$*

$$D_i^{\hat{\eta}}(\Gamma x) = 0 \Leftrightarrow \exists \alpha_i \geq 0 \text{ s.t. } (\Gamma x)_i = \alpha_i (\Gamma \hat{x})_i. \quad (39)$$

and we have for $i \in \hat{\mathcal{I}}^c$

$$D_i^{\hat{\eta}}(\Gamma x) = 0 \Leftrightarrow (\Gamma x)_i = 0_b \text{ or } \hat{\eta}_i = \frac{(\Gamma x)_i}{\|(\Gamma x)_i\|_2}. \quad (40)$$

As noticed in the image fusion model in [2], minimizing $D_i^{\hat{\eta}}(\Gamma x)$ enforces the alignment of the direction $(\Gamma x)_i$ with $\hat{\eta}_i$ and it is an efficient way to avoid contrast inversion.

In the following sections, unless stated otherwise, we will always consider $\hat{\eta} \in \Delta(\hat{x})$ and $\hat{p} = \Gamma^t \hat{\eta} \in \Omega(\hat{x})$.

4.1.2 Iterative Bregman regularization

The Bregman process [29] reduces the bias of solutions of (1) by successively solving problems of the form

$$\tilde{x}_{l+1} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda D_{\|\Gamma \cdot\|_{1,2}}^{\tilde{p}_l}(x, \tilde{x}_l) . \quad (41)$$

with $\tilde{p}_l \in \Omega(\tilde{x}_l)$. We consider a fixed λ , but different strategies can be considered with decreasing parameters λ_l as in [36, 38]. For $l = 0$, setting $\tilde{x}_0 = 0_n$, and taking $\tilde{p}_0 = 0_n \in \Omega(\tilde{x}_0)$ so that $D_{\|\Gamma \cdot\|_{1,2}}^{\tilde{p}_0}(x, \tilde{x}_0) = \|\Gamma x\|_{1,2}$, the first step exactly gives the biased solution of (1) with $\tilde{x}_1 = \hat{x}$. We denote by $\tilde{x}^{\text{IB}(\hat{p})} = \tilde{x}_2$ the solution obtained after 2 steps of the Iterative Bregman (IB) procedure (41):

$$\tilde{x}^{\text{IB}(\hat{p})} = \tilde{x}_2 \in \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda D_{\|\Gamma \cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) . \quad (42)$$

As underlined in relation (39), by minimizing $D_{\|\Gamma \cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) = \sum_{i=1}^m D_i^{\hat{\eta}}(\Gamma x)$, one aims at preserving the direction of $\Gamma \hat{x}$ on the support $\hat{\mathcal{L}}$, without ensuring $\operatorname{supp}(\Gamma \tilde{x}^{\text{IB}(\hat{p})}) \subseteq \hat{\mathcal{L}}$.

For the iterative framework, the support of the previous solution may indeed not be preserved ($\|(\Gamma \tilde{x}_l)_i\|_2 = 0 \not\Rightarrow \|(\Gamma \tilde{x}_{l+1})_i\|_2 = 0$) and can hence grow. The support of Γx_0 for $\tilde{x}_0 = 0_n$ is for instance totally empty whereas the one of $\hat{x} = \tilde{x}_1$ may not (and should not) be empty. For $l \rightarrow \infty$, the process actually converges to some x such that $\Phi x = y$. Because the IB procedure does not preserve the support of the solution, it cannot be considered as a refitting procedure and is more related to boosting approaches as discussed in Section 1.

4.1.3 Bregman based refitting

In order to respect the support of the biased solution \hat{x} and to keep track of the direction $\Gamma \hat{x}$ during the refitting, the authors of [5] proposed the following model:

$$\tilde{x}^{\text{B}(\hat{p})} \in \operatorname{argmin}_{x; \hat{p} \in \Omega(x)} \frac{1}{2} \|\Phi x - y\|_2^2 . \quad (43)$$

This model enforces the Bregman divergence to be 0, since, from eq. (29), we have:

$$\hat{p} \in \Omega(x) \Rightarrow D_{\|\Gamma \cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) = 0 . \quad (44)$$

We see from (39) that for $i \in \hat{\mathcal{L}}$, the direction of $(\Gamma \hat{x})_i$ is preserved in the refitted solution. From (40), we also observe that the absence of support is also preserved for any $i \in \hat{\mathcal{L}}^c$ where $\|\hat{\eta}_i\|_2 < 1$. Note that extra elements in the support $\Gamma \tilde{x}^{\text{B}(\hat{p})}$ may be added at coordinates $i \in \hat{\mathcal{L}}^c$ where $\|\hat{\eta}_i\|_2 = 1$.

4.1.4 Infimal Convolutions of Bregman (ICB) distances based refitting

To get rid of the direction dependency, the ICB (Infimal Convolutions of Bregman distances) model is also proposed in [5]:

$$\tilde{x}^{\text{ICB}(\hat{p})} \in \operatorname{argmin}_{x; \pm \hat{p} \in \Omega(x)} \frac{1}{2} \|\Phi x - y\|_2^2 . \quad (45)$$

The orientation model may nevertheless involve contrast inversions between biased and refitted solutions. In practice, relaxations are used in [5] by solving, for a large value $\gamma > 0$,

$$\tilde{x}_\gamma^{\text{B}(\hat{p})} \in \operatorname{argmin}_{x \in \mathbb{R}^n} \frac{1}{2} \|\Phi x - y\|_2^2 + \gamma D_{\|\Gamma \cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) . \quad (46)$$

The main advantage of this refitting strategy is that no support identification is required since everything is implicitly encoded in the subgradient \hat{p} . This makes the process stable even if the estimation of \hat{x} is not highly accurate. The support of $\Gamma\hat{x}$ is nevertheless only approximately preserved, since the constraint $D_{\|\Gamma\cdot\|_{1,2}}^{\hat{p}}(x, \hat{x}) = 0$ can never be ensured numerically with a finite value of γ .

4.1.5 The best of both Bregman worlds

From relations (37) and (38), the refitting models given in (42) and (43) can be reexpressed as a function of $\hat{\eta}$

$$\tilde{x}^{\text{IB}(\hat{\eta})} \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \sum_{i=1}^m D_i^{\hat{\eta}}(\Gamma x) , \quad (47)$$

$$\tilde{x}^{\text{B}(\hat{\eta})} \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \|\Phi x - y\|_2^2 \text{ s.t. } D_i^{\hat{\eta}}(\Gamma x) = 0, \forall i \in [m] . \quad (48)$$

Alternatively we can defined a mixed model, that we coin Best of Both Bregman (BBB), as

$$\tilde{x}^{\text{BBB}(\hat{\eta})} \in \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \frac{1}{2} \|\Phi x - y\|_2^2 + \lambda \sum_{i \in \hat{\mathcal{I}}} D_i^{\hat{\eta}}(\Gamma x) , \quad (49)$$

$$\text{s.t. } D_i^{\hat{\eta}}(\Gamma x) = 0, \forall i \in \hat{\mathcal{I}}^c . \quad (50)$$

With such reformulations, connections between refitting models (47), (48) and (49) can be clarified. The solution \tilde{x}^{IB} [29] is too relaxed, as it only penalizes the directions $(\Gamma x)_i$ using $(\Gamma \hat{x})_i$, without aiming at preserving the support of \hat{x} . The solution \tilde{x}^{B} [5] is too constrained: the direction within the support is required to be preserved exactly. The proposed refitting \tilde{x}^{BBB} lies in-between: it preserves the support, while authorizing some directional flexibility, as illustrated by the sharper square edges in Figure 1(g).

An important difference with BBB is that we consider local inclusions of subgradients of the function $\lambda \|\cdot\|_{1,2}$ at point $(\Gamma x)_i$ instead of the global inclusion of subgradients of the function $\lambda \|\Gamma \cdot\|_{1,2}$ at point x as in (43) and (45). Such a change of paradigm allows to adapt the refitting locally by preserving the support while including the flexibility of the original Bregman approach [29].

4.2 Covariant LEAst Square Refitting (CLEAR)

We now describe an alternative way for performing variational refitting. When specialized to $\ell_{1,2}$ sparse analysis regularization, CLEAR, a general refitting framework [16], consists in computing

$$\begin{aligned} \tilde{x}^{\text{CLEAR}} \in & \underset{x; \operatorname{supp}(\Gamma x) \subseteq \hat{\mathcal{I}}}{\operatorname{argmin}} \frac{1}{2} \|\Phi x - y\|_2^2 \\ & + \sum_{i \in \hat{\mathcal{I}}} \frac{\lambda}{2 \|(\Gamma \hat{x})_i\|_2} \left\| (\Gamma x)_i - P_{(\Gamma \hat{x})_i}((\Gamma x)_i) \right\|^2 , \end{aligned} \quad (51)$$

where we recall that $P_{(\Gamma \hat{x})_i}(\cdot)$ is the orthogonal projection onto $\operatorname{Span}(\Gamma \hat{x})_i$. This model promotes refitted solutions preserving to some extent the orientation $\Gamma \hat{x}$ of the biased solution. It also shrinks the amplitude of Γx all the more that the amplitude of $\Gamma \hat{x}$ are small. This penalty does not promote any kind of direction preservation, and as for the ICB model, contrast inversions may be observed between biased and refitted solutions. The quadratic term also over-penalizes large changes of orientation.

4.3 Equivalence

In the next proposition we show that, for a given vector $\bar{\eta}$, the concurrent Bregman based refitting techniques coincides with two of the refitting block penalties introduced in Section 2, namely HD, HO, and that SD is nothing else than the proposed Best of Both Bregman based penalty.

Proposition 6. *Let $\bar{\eta} \in \mathbb{R}^m$ be defined as*

$$\bar{\eta} = \arg \min_{\eta \in \Delta(\hat{x})} \|\eta\|_2 . \quad (52)$$

Then, for any y , we have the following equalities

1. $\tilde{x}^{\text{B}(\bar{\eta})} = \tilde{x}^{\text{HD}}$,
2. $\tilde{x}^{\text{ICB}(\bar{\eta})} = \tilde{x}^{\text{HO}}$,
3. $\tilde{x}^{\text{BBB}(\bar{\eta})} = \tilde{x}^{\text{SD}}$.

where the equalities have to be understood as an equality of the corresponding sets of minimizers.

The proof of Proposition 6 can be found in Appendix A. Additionally, we can show that CLEAR corresponds to the QO refitting block penalty also introduced in Section 2.

Proposition 7. *For any y , we have $\tilde{x}^{\text{CLEAR}} = \tilde{x}^{\text{QO}}$ where the equality has to be understood as an equality of the sets of minimizers.*

We invite the reader to refer to Appendix A for a proof of Proposition 7.

5 Experiments and Results

5.1 Toy experiments with TViso

We first consider TViso regularization of grayscale images on denoising problems of the form $y = x + w$ where w is an additive white Gaussian noise with standard deviation σ . We start with a simple toy example of a 128×128 image² composed of elementary geometric shapes, and we chose $\sigma = 50$. The corresponding image and its noisy version are given in Figs. 2.(a) and (b) respectively. To highlight the different behaviors between the six refitting block penalties, we chose to set the regularisation parameter of TViso to a large value $\lambda = 750$ leading to a strong bias in the solution. Fig. 2.(c) depicts this solution in which not only some structures are lost (the darkest disk), but a large loss of contrast can be observed (the white large square became gray). Unlike boosting, the purpose of refitting is not to recover the lost structures, but only to recover the correct amplitudes of the reconstructed objects without changing their geometrical aspect. In a second scenario, we consider the standard *cameraman* image of size 256×256 , and we set $\sigma = 20$ and $\lambda = 36$. The corresponding images are given in Figs. 3.(a-c).

We applied the iterative primal-dual algorithm with our joint-refitting (Algorithm (20)) for 4,000 iterations, with $\tau = 1/\|\nabla\|_2$, $\kappa = 1/\|\nabla\|_2$ and $\theta = 1$, and where $\|\nabla\|_2 = 2\sqrt{2}$. Results of refitting with HO, HD, QO, QD, SO and SD are given for the two scenarios on Figs. 2.(d-i) and Figs. 3.(d-i), respectively. As a quantitative measure of refitting between an estimate \hat{x} and the underlying image x , we consider the Peak Signal to Noise Ratio (PSNR) defined in this case as

$$\text{PSNR} = -10 \log_{10} \frac{1}{n} \|x - \hat{x}\|_2^2 . \quad (53)$$

The higher the PSNR, the better the refitting is supposed to be. In Fig. 2, we first observe that block penalties that are only based on orientation (HO, QO and SO) lead to refitted solution

²In this paper, we always consider images whose values are in the range $[0, 255]$.

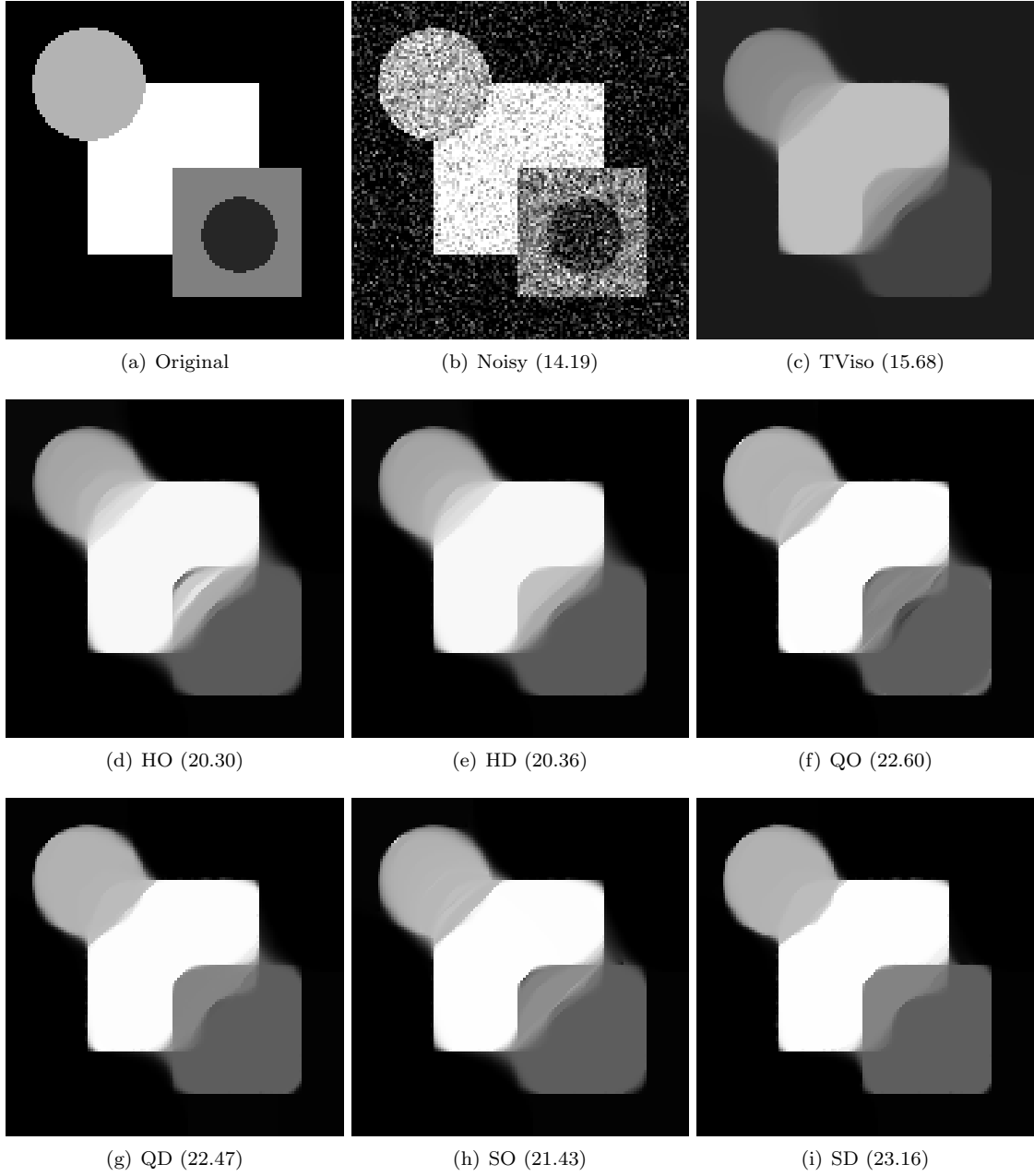


Figure 2: Comparison of alternative refitting approaches with the proposed SD model on a synthetic image.

with contrast inversions compared to the original solution of TViso (see the banana dark shapes at the interface between the two squares). In that scenario, HD, QD and SD provides satisfying visual quality, the latter achieving highest PSNR. In the context of Fig. 3, the image being more complex, the TViso solution presents more level lines than in the previous scenario. The block penalties HO and HD must preserve exactly these level lines. The refitted problem becomes too constrained and this lack of flexibility prevents HO and HD to deviate much from TViso. Their refitted solution remains significantly biased. Comparing now the quadratic penalties (QO



Figure 3: Comparison of alternative refitting approaches with the proposed SD model on the Cameraman image.

and QD) against the soft ones (SO and SD) reveals that quadratic penalties does not allow to enhance as much the contrast of some objects (see, *e.g.*, the second tower on the right of the image). Again SD achieved the highest PSNR in that second scenario.

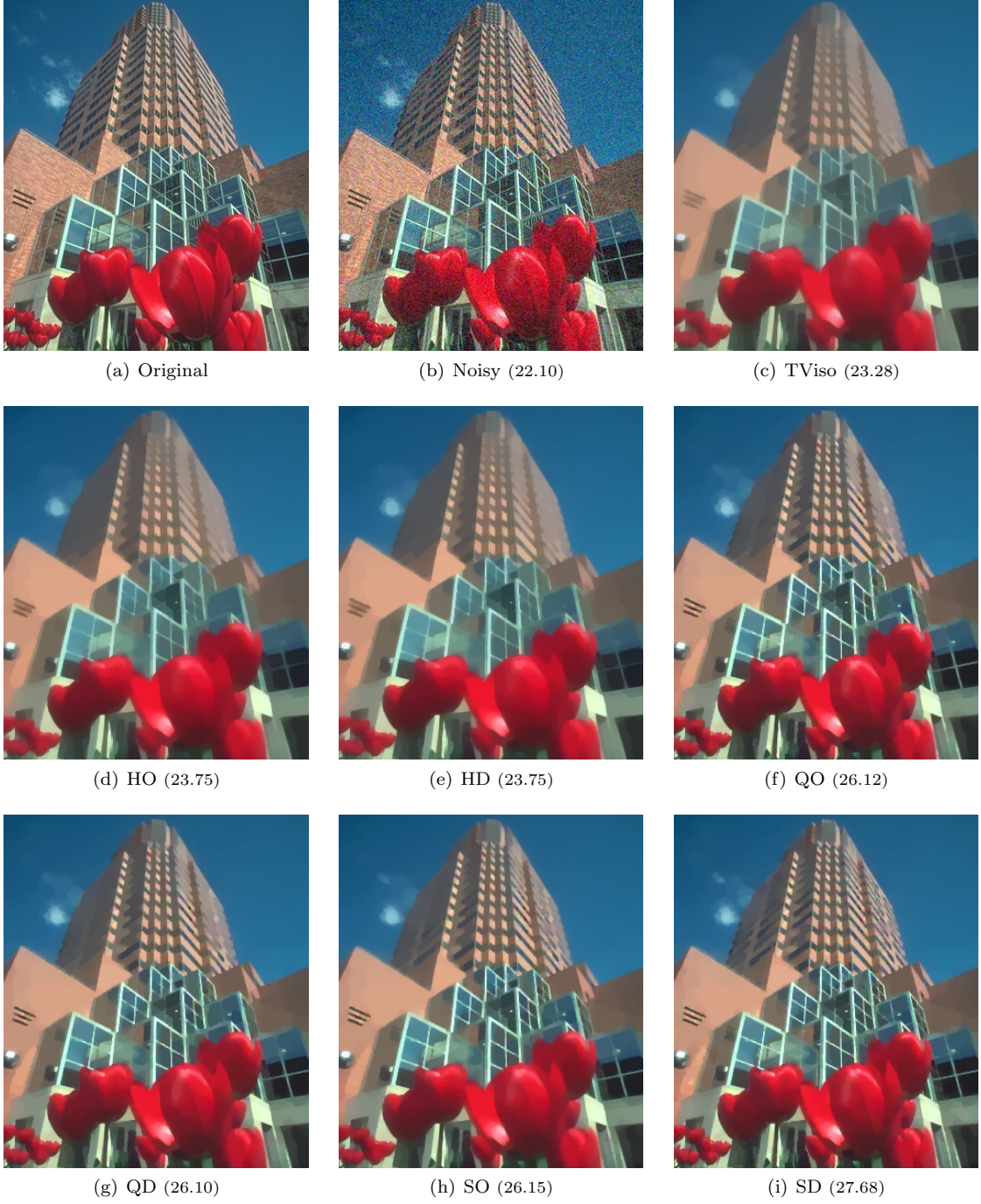


Figure 4: (a) A color image. (b) A corrupted version by Gaussian noise with standard deviation $\sigma = 20$. (b) Solution of TViso. Debiased solution with (d) HO, (e) HD, (f) QO, (g) QD, (h) SO and (i) SD. The Peak Signal to Noise Ratio (PSNR) is indicated in brackets bellow each image.

5.2 Experiments with color TViso

We now consider TViso regularization of degraded color images with three channels Red, Green, Blue (RGB). We defined blocks obtained by applying $\Gamma = (\nabla_R^1, \nabla_R^2, \nabla_G^1, \nabla_G^2, \nabla_B^1, \nabla_B^2)$, where

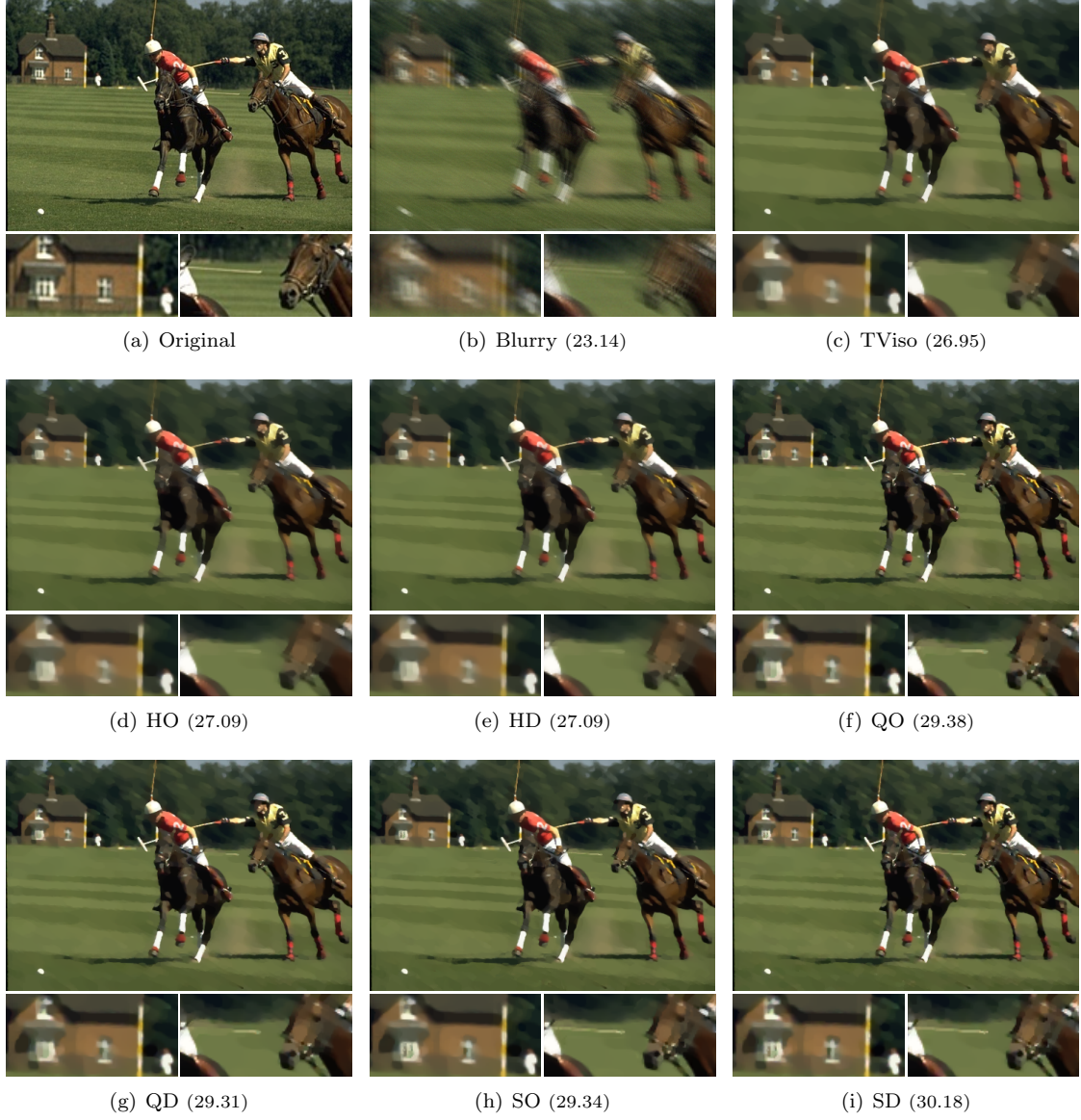


Figure 5: (a) A color image. (b) A corrupted version by a directional blur and Gaussian noise with standard deviation $\sigma = 2$. (c) Solution of TViso. Debiased solution with (d) HD, (e) QO and (f) SD. The PSNR is indicated in brackets below each image.

$m = n$, $b = 6$, and ∇_C^d denotes the forward discrete gradient in the direction $d \in \{1, 2\}$ for the color channel $C \in \{R, G, B\}$.

We first focused on a denoising problem $y = x + w$ where x is a color image and w is an additive white Gaussian noise with standard deviation $\sigma = 20$. We next focused on a deblurring problem $y = \Phi x + w$ where x is a color image, Φ is a convolution simulating a directional blur, and w is an additive white Gaussian noise with standard deviation $\sigma = 2$. In both cases, we chose $\lambda = 4.3\sigma$. We apply the iterative primal-dual algorithm with our joint-refitting (Algorithm (20)) for 1,000 iterations, with $\theta = 1$, $\tau = 1/\|\Gamma\|_2$ and $\kappa = 1/\|\Gamma\|_2$, where, as for the grayscale case, $\|\Gamma\|_2 = 2\sqrt{2}$. Note that in order to use the primal-dual algorithm, we have to implement

Γ^t , defined for $z = (z_R, z_G, z_B) \in \mathbb{R}^{n \times 6}$ where z_R, z_G, z_B are 2d vector fields, as given by

$$\Gamma^t z = \begin{pmatrix} -\operatorname{div} z_R & -\operatorname{div} z_G & -\operatorname{div} z_B \end{pmatrix}. \quad (54)$$

Results are provided on Fig. 4 and 5. Comparisons of refitting with our proposed SD block penalty, HO, HD, QO, QD and SO are provided. Using our proposed SD block penalty offers the best refitting performances in terms of both visual and quantitative measures. The loss of contrast of TViso is well-corrected, amplitudes are enhanced while smoothness and sharpness of TViso is preserved. Meanwhile, the approach does not create artifacts, invert contrasts, or reintroduce information that were not recovered by TViso.

5.3 Experiments with second order TGV

We finally consider the second order Total Generalized Variation (TGV) regularization [4] of grayscale images that can be expressed, for an image $x \in \mathbb{R}^n$ and regularization parameters $\lambda > 0$ and $\zeta \geq 0$, as

$$\begin{aligned} \hat{x} \in \operatorname{argmin}_{x \in \mathbb{R}^n} & \frac{1}{2} \|\Phi x - y\|_2^2 \\ & + \lambda \min_{z \in \mathbb{R}^{n \times 2}} \|\nabla x - \zeta z\|_{1,2} + \|\mathcal{E}(z)\|_{1,F} \end{aligned} \quad (55)$$

where $\nabla = (\nabla^1, \nabla^2) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times 2}$, ∇^d denotes the forward discrete gradient in the direction $d \in \{1, 2\}$, and $\mathcal{E} : \mathbb{R}^{n \times 2} \rightarrow \mathbb{R}^{n \times 2 \times 2}$ is a symmetric tensor field operator defined for a 2d vector field $z = (z^1, z^2)$ as

$$\mathcal{E}(z) = \begin{pmatrix} \bar{\nabla}^1 z^1 & \frac{1}{2}(\bar{\nabla}^1 z^2 + \bar{\nabla}^2 z^1) \\ \frac{1}{2}(\bar{\nabla}^1 z^2 + \bar{\nabla}^2 z^1) & \bar{\nabla}^2 z^2 \end{pmatrix} \quad (56)$$

where $\bar{\nabla}^d$ denotes the backward discrete gradient in the direction $d \in \{1, 2\}$, and $\|\cdot\|_{1,F}$ is the pointwise sum of the Frobenius norm of all matrices of a field. One can observe that for $\zeta = 0$ this model is equivalent to TViso. Interestingly the solution of the second order TGV can be obtained by solving a regularized least square problem with an ℓ_{12} sparse analysis term as

$$\hat{X} \in \operatorname{argmin}_{X \in \mathbb{R}^{n \times 3}} \frac{1}{2} \|\Xi X - y\|_2^2 + \lambda \|\Gamma X\|_{1,2} \quad (57)$$

where for an image $x \in \mathbb{R}^n$ and a vector field $z \in \mathbb{R}^{n \times 2}$, we consider $X = (x, z) \in \mathbb{R}^{n \times 3}$ and $\Xi : (x, z) \rightarrow \Phi x$ and we define $\Gamma : \mathbb{R}^{n \times 3} \mapsto \mathbb{R}^{2n \times 3}$ as

$$\Gamma(x, z) = \begin{pmatrix} \nabla^1 x - \zeta z^1 & \nabla^2 x - \zeta z^2 & 0 \\ \bar{\nabla}^1 z^1 & \bar{\nabla}^2 z^2 & \frac{1}{\sqrt{2}}(\bar{\nabla}^1 z^2 + \bar{\nabla}^2 z^1) \end{pmatrix}. \quad (58)$$

After solving (57), the solution of (55) is obtained as $\hat{x} = \Xi \hat{X}$. We use Chambolle Pock algorithm for which we also need to implement the adjoint $\Gamma^t : \mathbb{R}^{2n \times 3} \rightarrow \mathbb{R}^{n \times 3}$ of Γ given for fields $z \in \mathbb{R}^{n \times 2}$ and $e \in \mathbb{R}^{n \times 3}$ by

$$\Gamma^t \begin{pmatrix} z^1 & z^2 & \cdot \\ e^2 & e^2 & e^3 \end{pmatrix} = - \begin{pmatrix} (\bar{\nabla}^1 z^1 + \bar{\nabla}^2 z^2)^t \\ (\zeta z^1 + \nabla^1 e^1 + \frac{1}{\sqrt{2}} \nabla^2 e^3)^t \\ (\zeta z^2 + \nabla^1 e^2 + \frac{1}{\sqrt{2}} \nabla^2 e^3)^t \end{pmatrix}^t. \quad (59)$$

We focused on a denoising problem $y = x + w$ where x is a simulated elevation profile, ranging in $[0, 255]$, and w is an additive white Gaussian noise with standard deviation $\sigma = 2$. We chose $\lambda = 15$ and $\zeta = 0.45$. We applied the iterative primal-dual algorithm with our joint-refitting

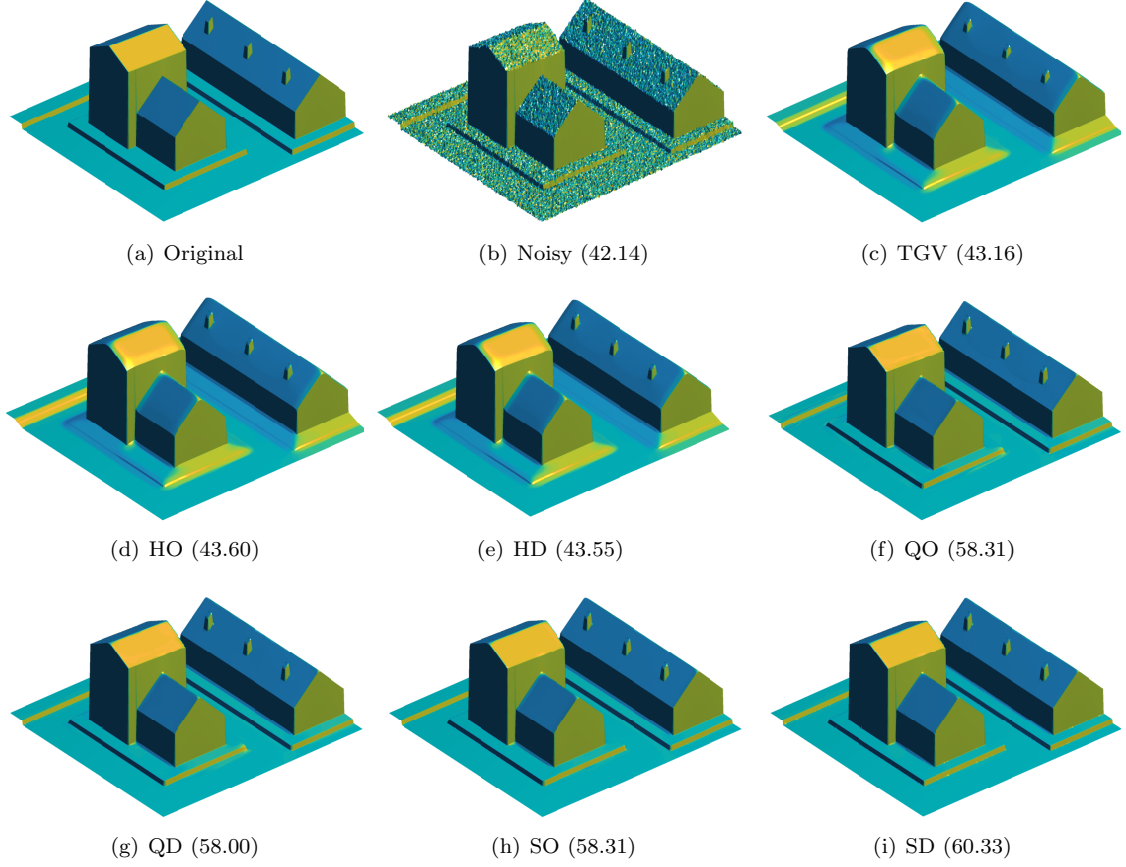


Figure 6: (a) An elevation profile (range 0 to 255). (b) A corrupted version by Gaussian noise with standard deviation $\sigma = 0.5$. (c) Solution of TGV. Debiased solution with (d) HO, (e) HD, (f) QO, (g) QD, (h) SO and (i) SD. The PSNR is indicated in brackets below each profile.

(Algorithm (20)) for 4,000 iterations, with $\tau = 1/\|\Gamma\|_2$, $\kappa = 1/\|\Gamma\|_2$ and $\theta = 1$, and where we estimated $\|\Gamma\|_2 \approx 3.05$ by power iteration (note that the quantity $\|\Gamma\|_2$ depends on the value of ζ).

Refitting comparisons between our proposed SD block penalty, and the five other alternatives are provided on Fig. 6. Using the SD model offers the best refitting performances in terms of both visual and quantitative measures. The bias of TGV is well-corrected, elevations (see the chimneys) and slopes (see the sidewalks) are enhanced while smoothness and sharpness of TGV is preserved. Meanwhile, the approach does not create artifacts, invert contrasts, or reintroduce information that was not recovered by TGV.

6 Conclusion

We have presented a block penalty formulation for the refitting of solutions obtained with ℓ_{12} sparse regularization. Through this framework, desirable properties of refitted solutions can easily be promoted. Different block penalty functions have been proposed and their properties discussed. We have namely introduced the SD block penalty that interpolates between Bregman iterations [29] and direction preservation [5].

Based on standard optimization schemes, we have also defined stable numerical strategies to jointly estimate a biased solution and its refitting. In order to take advantage of our efficient joint-refitting algorithm, we underline that it is important to consider *simple* block penalty functions, which proximal operator can be computed explicitly. This is the case for all the presented block penalties, more complex ones having been discarded from this paper for this reason.

Initially designed for TViso regularization, the approach has finally been extended to a generalized TGV model. Experiments show how the block penalties are able to preserve different structures of the biased solutions, while recovering the correct signal amplitude.

For the future, a challenging problem would be to define similar simple numerical schemes able to consider global properties on the refitting, such as the preservation of the tree of shapes [46] of the biased solution.

Acknowledgements This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 777826.

A Proof of Propositions

Proposition 1. • (a) As \tilde{x}^ϕ is solution of (6), the relation is obtained by observing that $\phi(\hat{z}, \hat{z}) = 0$ with (P1).

• (b) Looking at a ray $[0, z]$ for any vector z , this is a consequence of the convexity of ϕ and the fact that $\phi(z, \hat{z}) = 0$ for $\|z\|_2 = 0$.

• (c) Since $\cos(z, \hat{z}) = \cos(-z, \hat{z})$, one can combine $(z', z'') = (z, -z)$ and $(z', z'') = (-z, z)$ in (P2) to obtain (c).

• (d) Direct consequence of points (b) and (c).

• (e) This is due to the continuity of the function $\phi(\cdot, \hat{z})$ that is 0 on the ray $[0, \hat{z}]$ from (P1). \square

Proposition 2. We just give a sketch of the proof. More details can be found in [15]. On the support, one has $(\hat{\xi}_i^k, \hat{x}_i^k, \hat{v}_i^k) \rightarrow (\lambda \Gamma \hat{x}_i^k / \|(\Gamma \hat{x}^k)_i\|_2, \hat{x}_i^k, \hat{x}_i^k)$. Then for k sufficiently large, $\|\hat{\xi}_i^k + \kappa(\Gamma \hat{v}^k)_i\| \leq \lambda + \beta$ if and only if $i \in \mathcal{I}^c$. \square

Proposition 3. The method in [16] realizes the refitting through an algorithmic differentiation of the projection of the biased process. This leads to the Algorithm 20 except for the update of the dual variable $\tilde{\xi}_i$ on the support $i \in \hat{\mathcal{I}}^{k+1}$ that reads in [16]:

$$\tilde{\xi}_i^{k+1} = \frac{\lambda}{\|\hat{v}_i^{k+1}\|_2} \left(\hat{v}_i^{k+1} - P_{\hat{v}_i^{k+1}}(\tilde{v}_i^{k+1}) \right),$$

instead of

$$\tilde{\xi}_i^{k+1} = \text{prox}_{\kappa \phi_{\text{QO}}^*(\cdot, \Psi(\hat{v}_i^{k+1}), \hat{\mathcal{I}}^{k+1})}(\tilde{v}_i^{k+1})$$

in Algorithm (20). Using the proximal operator given by the QO block penalty in Table 2 gives : $\|\hat{v}_i^{k+1}\|_2 = \lambda + \kappa \|\Psi(\hat{v}_i^{k+1})\|_2$ which leads to the function (23) for estimating \hat{z}_i^k .

We deduce that $\Psi(\hat{v}_i^{k+1}) = \Psi(\hat{\xi}_i^k + \kappa \Gamma \hat{v}^k) \rightarrow (\Gamma \hat{x})_i$ from the convergence of the biased variables $(\hat{\xi}_i^k, \hat{x}_i^k, \hat{v}_i^k) \rightarrow (\lambda \Gamma \hat{x}_i^k / \|(\Gamma \hat{x}^k)_i\|_2, \hat{x}_i^k, \hat{x}_i^k)$. \square

Proposition 4. The proof is straightforward and reads as follow

$$D_{\|\cdot\|_{1,2}}^{\hat{\eta}}(x, \hat{x}) = \|\Gamma x\|_{1,2} - \langle \Gamma^t \hat{\eta}, x \rangle \quad (60)$$

$$= \|\Gamma x\|_{1,2} - \langle \hat{\eta}, \Gamma x \rangle \quad (61)$$

$$= \sum_{i=1}^m \underbrace{\|(\Gamma x)_i\|_2 - \langle \hat{\eta}_i, (\Gamma x)_i \rangle}_{D_{\|\cdot\|_2}^{\hat{\eta}_i}((\Gamma x)_i, (\Gamma \hat{x})_i)} . \quad (62)$$

□

Proposition 5. • For $i \in \hat{\mathcal{I}}$, i.e. $\|(\Gamma x)_i\|_2 > 0$, we get $\hat{\eta}_i = \frac{(\Gamma \hat{x})_i}{\|(\Gamma \hat{x})_i\|_2}$ from (36). We conclude from the definition $D_i^{\hat{\eta}}(\Gamma x) = \|(\Gamma x)_i\|_2 - \langle \hat{\eta}_i, (\Gamma x)_i \rangle = 0$.

• For $i \in \hat{\mathcal{I}}^c$ and using (36) we distinguish two cases from (31). If $\|\hat{\eta}_i\|_2 < 1$, then $D_i^{\hat{\eta}}(\Gamma x) = 0$ iff $(\Gamma x)_i = 0_b$. Otherwise one necessarily gets $\hat{\eta}_i = \frac{(\Gamma x)_i}{\|(\Gamma x)_i\|_2}$. □

Proposition 6. First notice that from (31), $\bar{\eta}$ can be written explicitly as

$$\bar{\eta}_i = \begin{cases} \frac{(\Gamma \hat{x})_i}{\|(\Gamma \hat{x})_i\|_2} & \text{if } i \in \hat{\mathcal{I}}, \\ 0_b & \text{otherwise,} \end{cases} \quad \text{for all } i \in [m]. \quad (63)$$

• [1.] This is a direct consequence of relation (44) and Proposition 5. Using (39), we have for $i \in \hat{\mathcal{I}}$ that $D_i^{\bar{\eta}}((\Gamma x)_i) = 0 \Leftrightarrow \cos((\Gamma x)_i, (\Gamma \hat{x})_i) = 1 \Leftrightarrow \cos((\Gamma x)_i, \bar{\eta}_i) = 1$. This is also valid for potential vanishing components $(\Gamma x)_i$ with the considered convention $\cos(0_b, \hat{z}) = 0$. We thus recover the penalty function HD in (11). Next, for all $i \in \hat{\mathcal{I}}^c$, we have $\bar{\eta}_i = 0$ by assumption, hence according to eq. (40):

$$D_i^{\bar{\eta}}(\Gamma x) = 0 \Leftrightarrow (\Gamma x)_i = 0_b \Leftrightarrow \text{supp}(\Gamma x) \subseteq \hat{\mathcal{I}}. \quad (64)$$

It follows that $\tilde{x}^{\text{B}(\bar{\eta})} = \tilde{x}^{\text{HD}}$. This case $\hat{\mathcal{I}}^c$ is the same for all next points.

• [2.] With the ICB model, we have $\pm \hat{p} \in \Omega(x)$. For $i \in \hat{\mathcal{I}}$, it gives with (39) that $D_i^{\pm \bar{\eta}}(\Gamma x) = 0 \Leftrightarrow \|(\Gamma x)_i\|_2 = \pm \langle (\Gamma x)_i, \hat{\eta}_i \rangle$. This is equivalent to $|\cos((\Gamma x)_i, (\Gamma \hat{x})_i)| = 1$ that corresponds to the penalty function HO in (10).

• [3.] For all $i \in \hat{\mathcal{I}}$, we have

$$D_i^{\hat{\eta}}((\Gamma x)_i) = \|(\Gamma x)_i\|_2 - \left\langle \frac{(\Gamma \hat{x})_i}{\|(\Gamma \hat{x})_i\|_2}, (\Gamma x)_i \right\rangle \quad (65)$$

$$= \|(\Gamma x)_i\|_2 (1 - \cos((\Gamma x)_i, (\Gamma \hat{x})_i)), \quad (66)$$

that gives the penalty function SD in (9). We then get that $\tilde{x}^{\text{BBB}(\bar{\eta})} = \tilde{x}^{\text{SD}}$. □

Proposition 7. The case $\hat{\mathcal{I}}^c$ is treated in Proposition 6. For $i \in \hat{\mathcal{I}}$ with the CLEAR model, we just have to observe that $\|(\Gamma x)_i - P_{(\Gamma \hat{x})_i}((\Gamma x)_i)\|^2 = \|(\Gamma x)_i\|_2^2 (1 - \cos^2((\Gamma x)_i, (\Gamma \hat{x})_i))$, which corresponds to the penalty function QO in (12). □

B Proximity operators of block penalties

B.1 Convex conjugates ϕ^*

We here compute the convex conjugate ϕ^* of the different block penalties $\phi(z, \hat{z})$ that only depends on $z \in \mathbb{R}^b$ and where $\hat{z} \in \mathbb{R}^b$ is a given fixed non null vector. We consider the following representation of the vectors z with respect to the \hat{z} axis: $z = \alpha \frac{\hat{z}}{\|\hat{z}\|_2} + \beta \frac{\hat{z}^\perp}{\|\hat{z}\|_2}$. This expression is valid for the case $b = 2$. If $b = 1$ then z is only parameterized by α . When $b > 2$, \hat{z}^\perp must be understood as a subspace S of dimension $b-1$ and β as a vector of $b-1$ components corresponding to each dimension of S . With this change of variables, we have $\|z\|_2^2 = \alpha^2 + \|\beta\|_2^2$ (since β is of dimension $b-1$), $\cos(z, \hat{z}) = \alpha / \sqrt{\alpha^2 + \|\beta\|_2^2}$, $P_{\hat{z}}(z) = \alpha \hat{z} / \|\hat{z}\|_2$ and $z - P_{\hat{z}}(z) = \beta \hat{z}^\perp / \|\hat{z}\|_2$. We

also observe for instance that $|\cos(z, \hat{z})| = 1 \Leftrightarrow \beta = 0_{b-1}$. All the block penalties $\phi(z, \hat{z})$ can thus be expressed as $\phi(\alpha, \beta)$. The convex conjugate reads

$$\phi^*(\alpha_0, \beta_0) = \sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \phi(\alpha, \beta) . \quad (67)$$

- [HO] It results that the convex conjugate $\phi_{\text{HO}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \iota_{\{0_{b-1}\}}(\beta) = \iota_{\{0\}}(\alpha_0) . \quad (68)$$

- [HD] The convex conjugate $\phi_{\text{HD}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \iota_{\mathbb{R}_+^* \times \{0_{b-1}\}}(\alpha, \beta) = \iota_{\mathbb{R}_-}(\alpha_0) . \quad (69)$$

- [QO] The convex conjugate $\phi_{\text{QO}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \frac{\lambda}{2} \frac{\alpha^2 + \|\beta\|_2^2}{\|\hat{z}\|_2} \left(1 - \frac{\alpha^2}{\alpha^2 + \|\beta\|_2^2} \right) \quad (70)$$

$$= \sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \frac{\lambda}{2} \frac{\|\beta\|_2^2}{\|\hat{z}\|_2} \quad (71)$$

$$= \begin{cases} \frac{\|\hat{z}\|_2 \|\beta_0\|_2^2}{2\lambda} & \text{if } \alpha_0 = 0 , \\ +\infty & \text{otherwise ,} \end{cases} \quad (72)$$

since the optimality condition on β gives us $\beta^* = \frac{\beta_0 \|\hat{z}\|_2}{\lambda}$.

- [QD] The convex conjugate $\phi_{\text{QD}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \frac{\lambda}{2} \frac{\|\beta\|_2^2}{\|\hat{z}\|_2} - \begin{cases} \frac{\lambda}{2} \frac{\alpha^2}{\|\hat{z}\|_2} & \text{if } \alpha \leq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (73)$$

$$= \begin{cases} \frac{\|\hat{z}\|_2 (\alpha_0^2 + \|\beta_0\|_2^2)}{2\lambda} & \text{if } \alpha_0 \leq 0 , \\ +\infty & \text{otherwise ,} \end{cases} \quad (74)$$

where we used that if $\alpha_0 > 0$, taking $\beta = 0$ and letting $\alpha \rightarrow \infty$ leads to $\phi_{\text{QD}}^*(\alpha_0, \beta_0) = +\infty$, otherwise, we used the optimality conditions on β and α giving us $\beta^* = \frac{\beta_0 \|\hat{z}\|_2}{\lambda}$ and $\alpha^* = \frac{\alpha_0 \|\hat{z}\|_2}{\lambda}$.

- [SO] The convex conjugate $\phi_{\text{SO}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \lambda \|\beta\|_2 = \iota_{\{0\} \times B_2^\lambda}(\alpha_0, \beta_0) , \quad (75)$$

since the optimality condition on β gives us $\|\beta_0\|_2 \leq \lambda$ if $\beta^* = 0$, and $\beta_0 = \lambda \frac{\beta^*}{\|\beta^*\|_2}$ otherwise. •

- [SD] The convex conjugate $\phi_{\text{SD}}^*(\alpha_0, \beta_0)$ is

$$\sup_{\alpha, \beta} \alpha_0 \alpha + \langle \beta_0, \beta \rangle - \frac{\lambda}{2} \left(\sqrt{\alpha^2 + \|\beta\|_2^2} - \alpha \right) \quad (76)$$

$$= \sup_{\alpha, \beta} (\alpha_0 + \lambda/2) \alpha + \langle \beta_0, \beta \rangle - \frac{\lambda}{2} \sqrt{\alpha^2 + \|\beta\|_2^2} \quad (77)$$

$$= \begin{cases} 0 & \text{if } \sqrt{\|\beta_0\|_2^2 + (\alpha_0 + \lambda/2)^2} \leq \lambda/2 , \\ +\infty & \text{otherwise ,} \end{cases} \quad (78)$$

since if $\sqrt{\|\beta_0\|_2^2 + (\alpha_0 + \lambda/2)^2} > \lambda/2$, then letting $\alpha \rightarrow \text{sign}(\alpha_0 + \lambda/2) \times \infty$ and $\beta \rightarrow \text{sign} \beta_0 \times \infty$ leads to $\phi_{\text{SD}}^*(\alpha_0, \beta_0) = +\infty$.

B.2 Computing $\text{prox}_{\kappa\phi^*}$

We here give the computation of the proximal operator $\text{prox}_{\kappa\phi^*}(\alpha_0, \beta_0)$ of the different ϕ^* that is given at point (α_0, β_0) by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \phi^*(\alpha, \beta). \quad (79)$$

- [HO] The proximal operator $\text{prox}_{\kappa\phi_{HO}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \iota_{\{0\}}(\alpha) = (0, \beta_0). \quad (80)$$

- [HD] The proximal operator $\text{prox}_{\kappa\phi_{HD}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \iota_{\mathbb{R}_-}(\alpha) = (\min(0, \alpha_0), \beta_0). \quad (81)$$

- [QO] The proximal operator $\text{prox}_{\kappa\phi_{QO}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \begin{cases} \frac{\|\hat{z}\|_2 \|\beta\|_2^2}{2\lambda} & \text{if } \alpha = 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (82)$$

$$= \frac{\lambda}{\lambda + \kappa \|\hat{z}\|_2} (0, \beta_0), \quad (83)$$

since the optimality condition gives $\lambda(\beta^* - \beta_0) + \kappa \|\hat{z}\|_2 \beta^* = 0$.

- [QD] The proximal operator $\text{prox}_{\kappa\phi_{QD}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \begin{cases} \frac{\|\hat{z}\|_2 (\alpha^2 + \|\beta\|_2^2)}{2\lambda} & \text{if } \alpha \leq 0, \\ +\infty & \text{otherwise.} \end{cases} \quad (84)$$

$$= \frac{\lambda}{\lambda + \kappa \|\hat{z}\|_2} (\min(0, \alpha_0), \beta_0).$$

- [SO] The proximal operator $\text{prox}_{\kappa\phi_{SO}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2 + \iota_{\{0\} \times B_2^\lambda}(\alpha_0, \beta_0)$$

$$= \frac{\lambda}{\max(\lambda, \|\beta_0\|_2)} (0, \beta_0). \quad (85)$$

- [SD] The proximal operator $\text{prox}_{\kappa\phi_{SD}^*}(\alpha_0, \beta_0)$ is given by

$$\underset{\alpha, \beta}{\operatorname{argmin}} \frac{1}{2\kappa} \left\| \begin{pmatrix} \alpha \\ \beta \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \right\|_2^2$$

$$+ \begin{cases} 0 & \text{if } \sqrt{\|\beta\|_2^2 + (\alpha + \lambda/2)^2} \leq \lambda/2, \\ +\infty & \text{otherwise.} \end{cases} \quad (86)$$

$$= \frac{\lambda}{2} \frac{(\alpha_0 + \lambda/2, \beta_0)}{\max(\lambda/2, \sqrt{\|\beta_0\|_2^2 + (\alpha_0 + \lambda/2)^2})} - (\lambda/2, 0), \quad (87)$$

which just corresponds to the projection of the ℓ_2 ball of \mathbb{R}^b of radius $\lambda/2$ and center $(-\lambda/2, 0)$.

References

- [1] Bach, F., Jenatton, R., Mairal, J., Obozinski, G., et al.: Optimization with sparsity-inducing penalties. *Foundations and Trends® in Machine Learning* **4**(1), 1–106 (2012)
- [2] Ballester, C., Caselles, V., Igual, L., Verdera, J., Rougé, B.: A variational model for p+ xs image fusion. *International Journal of Computer Vision* **69**(1), 43–58 (2006)
- [3] Belloni, A., Chernozhukov, V.: Least squares after model selection in high-dimensional sparse models. *Bernoulli* **19**(2), 521–547 (2013)
- [4] Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. *SIAM Journal on Imaging Sciences* **3**(3), 492–526 (2010)
- [5] Brinkmann, E.M., Burger, M., Rasch, J., Sutour, C.: Bias-reduction in variational regularization. *J. of Math. Imaging and Vision* (2017)
- [6] Bühlmann, P., Yu, B.: Boosting with the L2 loss: regression and classification. *J. Am. Statist. Assoc.* **98**(462), 324–339 (2003)
- [7] Burger, M., Gilboa, G., Moeller, M., Eckardt, L., Cremers, D.: Spectral decompositions using one-homogeneous functionals. *SIAM J. on Imaging Sciences* **9**(3), 1374–1408 (2016)
- [8] Burger, M., Korolev, Y., Rasch, J.: Convergence rates and structure of solutions of inverse problems with imperfect forward models. *Inverse Problems* **35**(2), 024006 (2019)
- [9] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. *J. of Math. Imaging and Vision* **40**, 120–145 (2011)
- [10] Chan, T., Marquina, A., Mulet, P.: High-order total variation-based image restoration. *SIAM Journal on Scientific Computing* **22**(2), 503–516 (2000)
- [11] Charest, M.R., Milanfar, P.: On iterative regularization and its application. *IEEE Transactions on Circuits And Systems For Video Technology* **18**(3), 406–411 (2008)
- [12] Chzhen, E., Hebiri, M., Salmon, J.: On lasso refitting strategies. *Bernoulli* **25**(4A), 3175–3200 (2019)
- [13] Combettes, P.L., Pesquet, J.C.: A douglas–rachford splitting approach to nonsmooth convex variational signal recovery. *Selected Topics in Signal Processing, IEEE Journal of* **1**(4), 564–574 (2007)
- [14] Deledalle, C.A., Papadakis, N., Salmon, J.: Contrast re-enhancement of total-variation regularization jointly with the douglas-rachford iterations. In: *Signal Processing with Adaptive Sparse Structured Representations* (2015)
- [15] Deledalle, C.A., Papadakis, N., Salmon, J.: On debiasing restoration algorithms: applications to total-variation and nonlocal-means. In: *Int. Conf. on Scale Space and Variational Methods in Computer Vision*, pp. 129–141. Springer (2015)
- [16] Deledalle, C.A., Papadakis, N., Salmon, J., Vaiter, S.: CLEAR: Covariant least-square refitting with applications to image restoration. *SIAM J. on Imaging Sciences* **10**(1), 243–284 (2017)
- [17] Deledalle, C.A., Papadakis, N., Salmon, J., Vaiter, S.: Refitting solutions promoted by ℓ_{12} sparse analysis regularizations with block penalties. In: *International Conference on Scale Space and Variational Methods in Computer Vision*, pp. 131–143. Springer (2019)

- [18] Dobson, D.C., Santosa, F.: Recovery of blocky images from noisy and blurred data. *SIAM Journal on Applied Mathematics* **56**(4), 1181–1198 (1996)
- [19] Douglas, J., Rachford, H.: On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society* pp. 421–439 (1956)
- [20] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. *Ann. Statist.* **32**(2), 407–499 (2004)
- [21] Elad, M., Milanfar, P., Rubinstein, R.: Analysis versus synthesis in signal priors. *Inverse problems* **23**(3), 947–968 (2007)
- [22] Esedoğlu, S., Osher, S.J.: Decomposition of images by the anisotropic rudin-osher-fatemi model. *Commun. Pure Appl. Math.* **57**(12), 1609–1626 (2004)
- [23] Gilboa, G.: A total variation spectral framework for scale and texture analysis. *SIAM J. Imaging Sci.* **7**(4), 1937–1961 (2014)
- [24] Lederer, J.: Trust, but verify: benefits and pitfalls of least-squares refitting in high dimensions. *arXiv preprint arXiv:1306.0113* (2013)
- [25] Lin, Y., Zhang, H.H.: Component selection and smoothing in multivariate nonparametric regression. *Ann. Statist.* **34**(5), 2272–2297 (2006)
- [26] Liu, J., Huang, T.Z., Selesnick, I.W., Lv, X.G., Chen, P.Y.: Image restoration using total variation with overlapping group sparsity. *Information Sciences* **295**, 232–246 (2015)
- [27] Milanfar, P.: A tour of modern image filtering: New insights and methods, both practical and theoretical. *IEEE Signal Processing Magazine* **30**(1), 106–128 (2013)
- [28] Nam, S., Davies, M., Elad, M., Gribonval, R.: The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.* **34**, 30–56 (2013)
- [29] Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterative regularization method for total variation-based image restoration. *Multiscale Model. & Simul.* **4**(2), 460–489 (2005)
- [30] Peyré, G., Fadili, J.: Group sparsity with overlapping partition functions. In: 2011 19th European Signal Processing Conference, pp. 303–307. *IEEE* (2011)
- [31] Peyré, G., Fadili, J., Chesneau, C.: Adaptive structured block sparsity via dyadic partitioning. In: *Proc. EUSIPCO 2011*, pp. 1455–1459 (2011)
- [32] Pierre, F., Aujol, J.F., Deledalle, C.A., Papadakis, N.: Luminance-guided chrominance denoising with debiased coupled total variation. In: *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pp. 235–248. *Springer* (2017)
- [33] Rigollet, P., Tsybakov, A.B.: Exponential screening and optimal rates of sparse estimation. *Ann. Statist.* **39**(2), 731–471 (2011)
- [34] Romano, Y., Elad, M.: Boosting of image denoising algorithms. *SIAM J. Imaging Sci.* **8**(2), 1187–1219 (2015)
- [35] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena* **60**(1-4), 259–268 (1992)

- [36] Scherzer, O., Groetsch, C.: Inverse scale space theory for inverse problems. In: M. Kerckhove (ed.) *Scale-Space and Morphology in Computer Vision*, pp. 317–325. Springer Berlin Heidelberg (2001)
- [37] Strong, D., Chan, T.: Edge-preserving and scale-dependent properties of total variation regularization. *Inverse problems* **19**(6), 165–187 (2003)
- [38] Tadmor, E., Nezzar, S., Vese, L.: A multiscale image representation using hierarchical (BV,L2) decompositions. *Multiscale Model. & Simul.* **2**(4), 554–579 (2004)
- [39] Talebi, H., Zhu, X., Milanfar, P.: How to saif-ly boost denoising performance. *IEEE Transactions on Image Processing* **22**(4), 1470–1485 (2013)
- [40] Tibshirani, R.: Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* pp. 267–288 (1996)
- [41] Tukey, J.W.: *Exploratory data analysis*. Reading, Mass. (1977)
- [42] Vaiter, S., Deledalle, C., Fadili, J., Peyré, G., Dossal, C.: The degrees of freedom of partly smooth regularizers. *Annals of the Institute of Statistical Mathematics* **69**(4), 791–832 (2017)
- [43] Vaiter, S., Deledalle, C.A., Peyré, G., Dossal, C., Fadili, J.: Local behavior of sparse analysis regularization: Applications to risk estimation. *Applied and Computational Harmonic Analysis* **35**(3), 433–451 (2013)
- [44] Vaiter, S., Peyré, G., Dossal, C., Fadili, J.: Robust sparse analysis regularization. *IEEE Transactions on Information Theory* **59**(4), 2001–2016 (2012)
- [45] Vaiter, S., Peyré, G., Fadili, J.: Model consistency of partly smooth regularizers. *IEEE Transactions on Information Theory* **64**(3), 1725–1737 (2017)
- [46] Weiss, P., Escande, P., Bathie, G., Dong, Y.: Contrast invariant SNR and isotonic regressions. *International Journal of Computer Vision* pp. 1–18 (2019)
- [47] Yu, G., Mallat, S., Bacry, E.: Audio denoising by time-frequency block thresholding. *IEEE Transactions on Signal processing* **56**(5), 1830–1839 (2008)
- [48] Yuan, M., Lin, Y.: Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68**(1), 49–67 (2006)