On debiasing restoration algorithms: applications to total-variation and nonlocal-means

Charles-Alban Deledalle, Nicolas Papadakis, Joseph Salmon

HAL Id: hal-01123542

https://hal.science/hal-01123542

Submitted on 5 Mar 2015

On debiasing restoration algorithms: applications to total-variation and nonlocal-means

Charles-Alban Deledalle[1,2], Nicolas Papadakis[1,2], Joseph Salmon[3]

[1] Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.
[2] CNRS, IMB, UMR 5251, F-33400 Talence, France.
{charles-alban.deledalle,nicolas.papadakis}@math.u-bordeaux.fr
[3] Institut Mines-Télécom, Télécom ParisTech, CNRS LTCI, Paris, France.
joseph.salmon@telecom-paristech.fr

Abstract. Bias in image restoration algorithms can hamper further analysis, typically when the intensities have a physical meaning of interest, e.g., in medical imaging. We propose to suppress a part of the bias – *the method bias* – while leaving unchanged the other unavoidable part – *the model bias*. Our debiasing technique can be used for any locally affine estimator including ℓ_1 regularization, anisotropic total-variation and some nonlocal filters.

1 Introduction

Restoration of an image of interest from its single noisy degraded observation necessarily requires imposing some regularity or *prior* on the solution. Being often only crude approximations of the true underlying signal of interest, such techniques always introduce a bias towards the *prior*. However, in general, this is not the only source of bias. In many cases, even though the model was perfectly accurate, the method would remain biased. This part of the bias often emerges from technical reasons, e.g., when approaching an NP-hard problem by an easier one (typically, using the ℓ_1 convex relaxation of an ℓ_0 pseudo-norm).

It is well known that reducing bias is not always favorable in terms of mean square error because of the so-called bias-variance trade-off. It is important to highlight that a debiasing procedure is expected to re-inject part of the variance, therefore increasing the residual noise. Hence, the mean square error is not always expected to be improved by such techniques. Debiasing is nevertheless essential in applications where the image intensities have a physical sense and critical decisions are taken from their values. For instance, the authors of [7] suggest using image restoration techniques to estimate a temperature map within a tumor tissue for real time automatic surgical intervention. In such applications, it is so crucial that the estimated temperature is not biased. A remaining residual noise is indeed favorable compared to an uncontrolled bias.

We introduce a debiasing technique that suppresses the extra bias – *the method bias* – emerging from the choice of the method and leave unchanged the bias that is due to the unavoidable choice of the model – *the model bias*. To

that end, we rely on the notion of model subspace essential to carefully define different notions of bias. This leads to a mathematical definition of debiasing for any locally affine estimators that respect some mild assumptions.

Interestingly, our debiasing definition for the ℓ_1 synthesis (also known as LASSO [20] or Basis Pursuit [5]) recovers a well known debiasing scheme called refitting that goes back to the "Hybrid LASSO" [9] (see [15] for more details).

For the ℓ_1 analysis [10], including the ℓ_1 synthesis but also the anisotropic total-variation [18], we show that debiasing can be performed with the same complexity as the primal-dual algorithm of [4] producing the biased estimate.

In other cases, e.g., for an affine version of the popular nonlocal-means [2], we introduce an iterative scheme that requires only a few run of an algorithm of the same complexity as the original one producing the biased estimate.

2   Background

We consider observing $f = f_0 + w \in \mathbb{R}^P$ a corrupted linear observation of an unknown signal $u_0 \in \mathbb{R}^N$ such that $f_0 = \Phi u_0$ where $\Phi \in \mathbb{R}^{N \times P}$ is a linear operator and w is a random vector modeling the noise fluctuations. We assume that $\mathbb{E}[w] = 0$ where $\mathbb{E}$ is the expectation operator. The linear operator Φ is a degrading operator typically with $P \leqslant N$ and with a non-empty kernel encoding some information loss such that the problem becomes ill-posed.

We focus on estimating the unknown signal u_0. Due to the ill-posedness of the observation model, we consider variational approaches that attempt to recover u_0 from the single observation f as a solution of the optimization problem

$$u_f^\star \in \underset{u \in \mathbb{R}^N}{\operatorname{argmin}} \, E(u, f) \, . \tag{1}$$

where $E : \mathbb{R}^N \times \mathbb{R}^P \to \mathbb{R}$ is assumed to have at least one minimum. The objective E is typically chosen to promote some structure, e.g., smoothness, piece-wise constantness, sparsity, etc., that is captured by the so-called *model subspace* $\mathcal{M}_f^\star$. Providing $u_f^\star$ is uniquely defined and differentiable at f, we define $\mathcal{M}_f^\star \subseteq \mathbb{R}^N$ as the tangent affine subspace at f of the mapping $f \mapsto u_f^\star$, i.e.,

$$\mathcal{M}_f^\star = u_f^\star + \operatorname{Im}[J_f^\star] = \left\{ u \in \mathbb{R}^N \, ; \, \exists z \in \mathbb{R}^P, u = u_f^\star + J_f^\star z \right\} \quad \text{with} \quad J_f^\star = \left. \frac{\partial u_f^\star}{\partial f} \right|_f \tag{2}$$

where $J_f^\star$ is the Jacobian operator at f of the mapping $f \mapsto u_f^\star$ (see [23] for an alternative but related definition of model subspace). When $u_f^\star \in \operatorname{Im}[J_f^\star]$, the model subspace restricts to the linear vector subspace $\mathcal{M}_f^\star = \operatorname{Im}[J_f^\star]$. In the rest of the paper, $u_f^\star$ is assumed to be differentiable at f_0 and for almost all f.

Example 1 *The least square estimator constrained to the affine subspace $C = b + \operatorname{Im}[A]$, $b \in \mathbb{R}^N$ and $A \in \mathbb{R}^{N \times Q}$, is a particular instance of* (1) *where*

$$E(u, f) = \|\Phi u - f\|^2 + \iota_C(u) \tag{3}$$

and for any set C, ι_C is its indicator function: $\iota_C(u) = 0$ if $u \in C$, $+\infty$ otherwise. The solution of minimum Euclidean norm is unique and given by

$$u_f^\star = b + A(\Phi A)^+ (f - \Phi b) \tag{4}$$

where for a matrix M, M^+ is its Moore-Penrose pseudo-inverse. The affine constrained least square restricts the solution $u_f^\star$ to the affine model subspace $\mathcal{M}_f^\star = b + \mathrm{Im}[A(\Phi A)^t]$ (as $\mathrm{Im}[M^+] = \mathrm{Im}[M^t]$). Taking $C = \mathbb{R}^N$ with for instance $Q = N$, $A = \mathrm{Id}$ and $b = 0$, leads to an unconstrained solution $u_f^\star = \Phi^+ f$ whose model subspace is $\mathcal{M}_f^\star = \mathrm{Im}[\Phi^t]$ reducing to $\mathbb{R}^N$ when Φ has full column rank.

Example 2 *The Tikhonov regularization (or Ridge regression) [21, 14] is another instance of* (1) *where, for some parameter $\lambda > 0$ and matrix $\Gamma \in \mathbb{R}^{L \times N}$,*

$$E(u, f) = \frac{1}{2}\|\Phi u - f\|^2 + \frac{\lambda}{2}\|\Gamma u\|^2 \ . \tag{5}$$

Provided $\mathrm{Ker}\, \Phi \cap \mathrm{Ker}\, \Gamma = \{0\}$, $u_f^\star$ is uniquely defined as $u_f^\star = (\Phi^t \Phi + \lambda \Gamma^t \Gamma)^{-1} \Phi^t f$ which has a linear model subspace given by $\mathcal{M}_f^\star = \mathrm{Im}[\Phi^t]$.

Example 3 *The hard thresholding [8], used when $\Phi = \mathrm{Id}$ and f_0 is supposed to be sparse, is a solution of* (1) *where, for some parameter $\lambda > 0$,*

$$E(u, f) = \frac{1}{2}\|u - f\|^2 + \frac{\lambda^2}{2}\|u\|_0 \ , \tag{6}$$

where $\|u\|_0 = \#\{i \in [P] \ ; \ u_i \neq 0\}$ counts the number of non-zero entries of u and $[P] = \{1, \ldots, P\}$. The hard thresholding operation writes

$$(u_f^\star)_{\mathcal{I}_f} = f_{\mathcal{I}_f} \quad and \quad (u_f^\star)_{\mathcal{I}_f^c} = 0 \tag{7}$$

where $\mathcal{I}_f = \{i \in [P] \ ; \ |f_i| > \lambda\}$ is the support of $u_f^\star$, $\mathcal{I}_f^c$ is the complement of $\mathcal{I}_f$ on $[P]$, and for any vector v, $v_{\mathcal{I}_f}$ is the sub-vector whose elements are indexed by $\mathcal{I}_f$. As $u_f^\star$ is piece-wise differentiable, its model subspace is only defined for almost all f as $\mathcal{M}_f^\star = \{u \in \mathbb{R}^N \ ; \ u_{\mathcal{I}_f^c} = 0\} = \mathrm{Im}[\mathrm{Id}_{\mathcal{I}_f}]$, where for any matrix M, $M_{\mathcal{I}_f}$ is the sub-matrix whose columns are indexed by $\mathcal{I}_f$. Note that $\mathrm{Id}_{\mathcal{I}_f} \in \mathbb{R}^{N \times \#\mathcal{I}_f}$.

Example 4 *The soft thresholding [8], used when $\Phi = \mathrm{Id}$ and f_0 is supposed to be sparse, is another particular solution of* (1) *where*

$$E(u, f) = \frac{1}{2}\|u - f\|^2 + \lambda\|u\|_1 \ , \tag{8}$$

with $\|u\|_1 = \sum_i |u_i|$ the ℓ_1 norm of u. The soft thresholding operation writes

$$(u_f^\star)_{\mathcal{I}_f} = f_{\mathcal{I}_f} - \lambda\,\mathrm{sign}(f_{\mathcal{I}_f}) \quad and \quad (u_f^\star)_{\mathcal{I}_f^c} = 0 \ , \tag{9}$$

where $\mathcal{I}_f$ is defined as above, and, as for the hard thresholding: $\mathcal{M}_f^\star = \mathrm{Im}[\mathrm{Id}_{\mathcal{I}_f}]$.

3   Bias of reconstruction algorithms

Due to the ill-posedness of our observation model and without any assumptions on u_0, one cannot ensure the noise variance to be reduced while keeping the solution $u_f^\star$ unbiased. Recall that the statistical bias is defined as the difference

$$\text{Statistical bias} = \mathbb{E}[u_f^\star] - u_0 \ . \tag{10}$$

An estimator is said unbiased when its statistical bias vanishes. Unfortunately the statistical bias is difficult to manipulate when $f \mapsto u_f^\star$ is non linear. We therefore restrict to a definition of bias at $f_0 = \Phi u_0$ as the error $u_{f_0}^\star - u_0$. Note that when $f \mapsto u_f^\star$ is affine, both definitions match (the expectation being linear). Most methods are biased since, without assumptions, u_0 cannot be guaranteed to be in complete accordance with the model subspace, i.e., $u_0 \notin \mathcal{M}_{f_0}^\star$. It is then important to distinguish techniques that are only biased due to a problem of modeling to the ones that are biased due to the method. We then define the model bias and the method bias as the quantities

$$u_{f_0}^\star - u_0 = \underbrace{u_{f_0}^\star - \Pi_{\mathcal{M}_{f_0}^\star}(u_0)}_{\text{Method bias}} - \underbrace{\Pi_{(\mathcal{M}_{f_0}^\star)^\perp}(u_0)}_{\text{Model bias}} \ , \tag{11}$$

where for any set S, Π_S denotes the orthogonal projection on S and $S^\perp$ denotes its orthogonal set. We now define a methodically unbiased estimator as follows.

Definition 1 *An estimator $u_f^\star$ is <u>methodically unbiased</u> if*

$$\forall u_0 \in \mathbb{R}^N, \quad u_{f_0}^\star = \Pi_{\mathcal{M}_{f_0}^\star}(u_0)$$

We also define the weaker concept of weakly unbiased estimator as follows.

Definition 2 *An estimator $u_f^\star$ is <u>weakly unbiased</u> if*

$$\forall u_0 \in \mathcal{M}_{f_0}^\star, \quad u_{f_0}^\star = u_0.$$

The quantity $u_{f_0}^\star - u_0$ for $u_0 \in \mathcal{M}_{f_0}^\star$ is called the weak bias of $u_f^\star$ at u_0.

Remark that a methodically unbiased estimator is also weakly unbiased.

Examples. The unconstrained least-square estimator is methodically unbiased since $u_{f_0}^\star = \Phi^+ f_0 = \Phi^+ \Phi u_0 = \Pi_{\text{Im}[\Phi^t]}(u_0) = \Pi_{\mathcal{M}_{f_0}^\star}(u_0)$. Moreover, being linear, it becomes statistically unbiased whenever Φ has full column rank since $\Phi^+\Phi =$ Id. However the constrained least-square estimator is only weakly unbiased: its methodical bias only vanishes when $u_0 \in \mathcal{M}_{f_0}^\star$, i.e., when there exists $t_0 \in \mathbb{R}^Q$ such that $u_0 = b + A(\Phi A)^t t_0$. The hard thresholding is also methodically unbiased remarking that $u_{f_0}^\star$ is the orthogonal projection on $\mathcal{M}_{f_0}^\star = \text{Im}[\text{Id}_{\mathcal{I}_{f_0}}]$. Unlike the unconstrained least-square estimator, Tikhonov regularization has a non zero weak bias. The soft thresholding is also known to be biased [11] and its weak bias is given by $-\lambda \text{Id}_{\mathcal{I}_{f_0}} \text{sign}(f_0)_{\mathcal{I}_{f_0}}$. Often, estimators are said to be unbiased when they are actually only weakly unbiased.

4 Definitions of debiasing

Given an estimate $u_f^\star$ of u_0, we define a debiasing of $u_f^\star$ as follows.

Definition 3 *An estimator $\tilde{u}_f^\star$ of u_0 is a <u>weak debiasing</u> of $u_f^\star$ if it is weakly unbiased and $\tilde{\mathcal{M}}_f^\star = \mathcal{M}_f^\star$ for almost all f, with $\tilde{\mathcal{M}}_f^\star$ the model subspace of $\tilde{u}_f^\star$ at f. Moreover, it is a <u>methodical debiasing</u> if it is also methodically unbiased.*

Examples. The unconstrained least square estimator is a methodical debiasing of the Tikhonov regularization, since it is a methodically unbiased estimator of u_0 and they share the same model subspace. The hard thresholding is a methodical debiasing of the soft thresholding, for the same reasons.

A good candidate for debiasing $u_f^\star$ is the constraint least squares on $\mathcal{M}_f^\star$:

$$\tilde{u}_f^\star = u_f^\star + U_f^\star(\Phi U_f^\star)^+(f - \Phi u_f^\star) \in \underset{u \in \mathcal{M}_f^\star}{\operatorname{argmin}} \|\Phi u - f\|^2 \tag{12}$$

where $U_f^\star \in \mathbb{R}^{N \times n}$ with $n = \operatorname{rank}[J_f^\star]$ is a matrix whose columns form a basis of $\operatorname{Im}[J_f^\star]$. Let $V_f^\star \in \mathbb{R}^{n \times P}$ be a matrix such that $J_f^\star = U_f^\star V_f^\star$. The following theorem shows that under mild assumptions this choice corresponds to a debiasing of $u_f^\star$.

Theorem 1 *Assume that $f \mapsto u_f^\star$ is locally affine for almost all f and that Φ is invertible on $\mathcal{M}_f^\star$. Then $\tilde{u}_f^\star$ defined in Eq. (12) is a weak debiasing of $u_f^\star$.*

Proof. Since $f \mapsto u_f^\star$ is locally affine, $f \mapsto U_f^\star$ can be chosen locally constant. Deriving (12) for almost all f leads to the Jacobian $\tilde{J}_f^\star$ of $\tilde{u}_f^\star$ given by

$$\tilde{J}_f^\star = \frac{\partial \tilde{u}_f^\star}{\partial f} = \frac{\partial u_f^\star}{\partial f} + U_f^\star(\Phi U_f^\star)^+ \left(\frac{\partial f}{\partial f} - \Phi \frac{\partial u_f^\star}{\partial f}\right) = J_f^\star + U_f^\star(\Phi U_f^\star)^+(\operatorname{Id} - \Phi J_f^\star)$$

$$= U_f^\star V_f^\star + U_f^\star(\Phi U_f^\star)^+(\operatorname{Id} - \Phi U_f^\star V_f^\star) = U_f^\star(\Phi U_f^\star)^+ , \tag{13}$$

since $\Phi U_f^\star$ has full column rank due to the assumption that Φ is invertible on $\mathcal{M}_f^\star$. It follows that

$$\tilde{\mathcal{M}}_f^\star = \tilde{u}_f^\star + \operatorname{Im}[\tilde{J}_f^\star] = u_f^\star + U_f^\star(\Phi U_f^\star)^+(f - \Phi u_f^\star) + \operatorname{Im}[U_f^\star(\Phi U_f^\star)^+] \tag{14}$$

$$= u_f^\star + \operatorname{Im}[U_f^\star(\Phi U_f^\star)^+] = u_f^\star + \operatorname{Im}[U_f^\star] = \mathcal{M}_f^\star , \tag{15}$$

since $\Phi U_f^\star$ has full column rank. Moreover, for any $u_0 \in \mathcal{M}_{f_0}^\star$, the equation $\Phi u = f_0$ has a unique solution $u = u_0$ in $\mathcal{M}_{f_0}^\star$ since Φ is invertible on $\mathcal{M}_{f_0}^\star$. Hence, $\tilde{u}_{f_0}^\star = u_0$ is the unique solution of (12), which concludes the proof. $\qquad\square$

The next proposition shows that the condition "Φ invertible on $\mathcal{M}_f^\star$" can be dropped when looking at $u_f^\star$ and $\tilde{u}_f^\star$ through Φ. The debiasing becomes furthermore methodical.

Proposition 1 *Assume $f \mapsto u_f^\star$ is locally affine for almost all f. Taking $\tilde{u}_f^\star$ defined in Eq. (12), then the predictor $\Phi \tilde{u}_f^\star$ of $f_0 = \Phi u_0$ is equal to $\Pi_{\Phi \mathcal{M}_f^\star}(f)$ and is a methodical debiasing of $\Phi u_f^\star$.*

Proof. Since $\Phi U_f^\star(\Phi U_f^\star)^+ = \Pi_{\operatorname{Im}[\Phi U_f^\star]}$, we have $\Phi \tilde{u}_f^\star = \Pi_{\Phi \mathcal{M}_f^\star}(f)$. As the orthogonal projector on its own model space, it is methodically unbiased. Moreover $\operatorname{Im}[\Pi_{\operatorname{Im}[\Phi U_f^\star]}] = \operatorname{Im}[\Phi U_f^\star]$, hence $\Phi \tilde{u}_f^\star$ and $\Phi u_f^\star$ share the same model subspace. $\square$

Remark 1 *As an immediate consequence, the debiasing of any locally affine denoising algorithm is a methodical debiasing, since $\Phi \tilde{u}_f^\star = \tilde{u}_f^\star$.*

We focus in the next sections on the debiasing of estimators without explicit expression for $\mathcal{M}_f^\star$, meaning that Eq. (12) cannot be used directly. We first introduce an algorithm for the case of ℓ_1 analysis relying on the computation of the directional derivative $J_f^\star f$. We propose next a general approach, applied to an affine nonlocal estimator, that requires $J_f^\star \delta$ for randomized directions δ.

5   Debiasing the ℓ_1 analysis minimization

From now on, the dependency of all quantities with respect to the observation f will be dropped for the sake of simplicity. Given a linear operator $\Gamma \in \mathbb{R}^{L \times N}$, the ℓ_1 analysis minimization reads, for $\lambda > 0$, as

$$E(u, f) = \frac{1}{2}\|\Phi u - f\|^2 + \lambda\|\Gamma u\|_1 \ . \tag{16}$$

Provided $\mathrm{Ker}\,\Phi \cap \mathrm{Ker}\,\Gamma = \{0\}$, there exists a solution given implicitly, see [22], as

$$u^\star = U(\Phi U)^+ f - \lambda U(U^t \Phi^t \Phi U)^{-1} U^t (\Gamma^t)_\mathcal{I} s_\mathcal{I} \tag{17}$$

for almost all f and where $\mathcal{I} = \{i \ ; \ (\Gamma u^\star)_i \neq 0\} \subseteq [L] = \{1, \ldots, L\}$ is called the co-support of the solution, $s = \mathrm{sign}(\Gamma u^\star)$, $U = U_f^\star$ is a matrix whose columns form a basis of $\mathrm{Ker}[\mathrm{Id}_{\mathcal{I}^c}^t \Gamma]$ and ΦU has full column rank. Note that $s_\mathcal{I}$ and U are locally constant almost everywhere since the co-support is stable with respect to small perturbations [22]. It then follows that the model subspace is implicitly defined as $\mathcal{M}^\star = \mathrm{Im}[U] = \mathrm{Ker}[\mathrm{Id}_{\mathcal{I}^c}^t \Gamma]$, and so, the ℓ_1 analysis minimization suffers from a weak bias equal to $-\lambda U(U^t \Phi^t \Phi U)^{-1} U^t (\Gamma^t)_\mathcal{I} s_\mathcal{I}$. Given that $u^\star \in \mathrm{Im}[U]$ and it is locally affine, its weak debiased solution is defined for almost all f as

$$\tilde{u}^\star = U(\Phi U)^+ f \ . \tag{18}$$

The ℓ_1 synthesis [20, 8] consists in taking $\Gamma = \mathrm{Id}$, hence $U = \mathrm{Id}_\mathcal{I}$, so (17) becomes

$$u_\mathcal{I}^\star = (\Phi_\mathcal{I})^+ f - \lambda((\Phi_\mathcal{I})^t \Phi_\mathcal{I})^{-1} s_\mathcal{I} \quad \text{and} \quad u_{(\mathcal{I})^c}^\star = 0 \ . \tag{19}$$

Its model subspace is implicitly defined as $\mathcal{M}^\star = \mathrm{Im}[\mathrm{Id}_\mathcal{I}]$, its weak bias is $-\lambda \mathrm{Id}_\mathcal{I}((\Phi_\mathcal{I})^t \Phi_\mathcal{I})^{-1} s_\mathcal{I}$ and its weak debiasing is $\tilde{u}^\star = \mathrm{Id}_\mathcal{I}(\Phi_\mathcal{I})^+ f$. Subsequently, taking $\Phi = \mathrm{Id}$ leads to the soft-thresholding presented earlier.

The anisotropic Total-Variation (TV) [18] is a particular instance of (16) where $u_0 \in \mathbb{R}^N$ can be identified to a d-dimensional discrete signal, for which $\Gamma \in \mathbb{R}^{L \times N}$, with $L = dN$, is the concatenation of the discrete gradient operators in each canonical directions. In this case $\mathcal{I}$ is the set of indexes where the solution has discontinuities (non-null gradients) and $\mathcal{M}^\star$ is the space of piece-wise constant signals sharing the same discontinuities as the solution. Its weak bias reveals a loss of contrast: a shift of intensity on each piece depending on its surrounding and the ratio between its perimeter and its area, as shown, e.g., in [19]. Note that the so-called *staircasing* effect of TV regularization is encoded in our framework as a model bias, and is therefore not reduced by our debiasing technique. Strategies devoted to the reduction of this effect have been studied in, e.g., [16].

Since in general $u^\star$ has no explicit solutions, it is usually estimated thanks to an iterative algorithm that can be expressed as a sequence u^k converging to $u^\star$. The question we address is how to compute $\tilde{u}^\star$ in practice, i.e., to evaluate Eq. (18), or more precisely, how to jointly build a sequence $\tilde{u}^k$ converging to $\tilde{u}^\star$.

We propose a technique that relies on the observation that, given (17), for almost all f, the Jacobian $J^\star$ of $u^\star$ at f applied to f, leads to Eq. (18), i.e.,

$$J^\star[f] = U(\Phi U)^+ f = \tilde{u}^\star \tag{20}$$

since U and $s_{\mathcal{I}}$ are locally constant [22]. We so define a sequence $\tilde{u}^k$ which is, up to a slight modification, the closed-form derivation of the primal-dual sequence u^k of [4]. Most importantly, we provide a proof of its convergence towards $\tilde{u}^\star$.

Note that other debiasing techniques could be employed for the ℓ_1 analysis, e.g., using iterative hard-thresholding [13, 1], refitting techniques [9, 15], post-refinement techniques based an Bregman divergences and nonlinear inverse scale spaces [17, 3, 24] or with ideal spectral filtering in the analysis sense [12].

5.1 Primal-dual algorithm

Before stating our main result, let us recall some of the properties of primal-dual techniques. Dualizing the ℓ_1 analysis norm $u \mapsto \lambda\|\Gamma u\|_1$, the primal problem can be reformulated as the following saddle-point problem

$$z^\star = \operatorname*{argmax}_{z \in \mathbb{R}^L} \ \min_{u \in \mathbb{R}^N} \frac{1}{2}\|\Phi u - f\|^2 + \langle \Gamma u, z \rangle - \iota_{B_\lambda}(z) \tag{21}$$

where $z^\star \in \mathbb{R}^L$ is the dual variable, and $B_\lambda = \{z \ ; \ \|z\|_\infty \leqslant \lambda\}$ is the ℓ_∞ ball.

First order primal-dual optimization. Taking $\sigma\tau < \frac{1}{\|\Gamma\|_2^2}$, $\theta \in [0, 1]$ and initializing (for instance,) $u^0 = v^0 = 0 \in \mathbb{R}^N$, $z^0 = 0 \in \mathbb{R}^L$, the primal-dual algorithm of [4] applied to problem (21) reads

$$\begin{cases} z^{k+1} = \Pi_{B_\lambda}(z^k + \sigma\Gamma v^k), \\ u^{k+1} = (\mathrm{Id} + \tau\Phi^t\Phi)^{-1}\left(u^k + \tau(\Phi^t f - \Gamma^t(z^{k+1}))\right), \\ v^{k+1} = u^{k+1} + \theta(u^{k+1} - u^k), \end{cases} \tag{22}$$

where the projection of z over B_λ is done component-wise as

$$\Pi_{B_\lambda}(z)_i = \begin{cases} z_i & \text{if} \ \ |z_i| \leqslant \lambda, \\ \lambda \operatorname{sign}(z_i) & \text{otherwise.} \end{cases} \tag{23}$$

The primal-dual sequence u^k converges to a solution $u^\star$ of (16) [4]. We assumed here that $u^\star$ verifies (17) with ΦU full-column rank. This could be enforced as shown in [22], but it did not seem to be necessary in our experiments.

5.2 Debiasing algorithm

As pointed out earlier, the debiasing of $u^\star$ consists in applying the Jacobian matrix $J^\star$ at f to f itself. This idea leads to the proposed debiasing algorithm

that constructs a sequence of debiased iterates from the original biased primal-dual sequence with initialization $\tilde{u}^0 = \tilde{v}^0 = 0 \in \mathbb{R}^N$, $\tilde{z}^0 = 0 \in \mathbb{R}^L$ as follows

$$\begin{cases} \tilde{z}^{k+1} = \Pi_{z^k + \sigma \Gamma v^k}(\tilde{z}^k + \sigma \Gamma \tilde{v}^k), \\ \tilde{u}^{k+1} = (\mathrm{Id} + \tau \Phi^t \Phi)^{-1} \left(\tilde{u}^k + \tau (\Phi^t f - \Gamma^t \tilde{z}^{k+1}) \right), \\ \tilde{v}^{k+1} = \tilde{u}^{k+1} + \theta(\tilde{u}^{k+1} - \tilde{u}^k), \end{cases} \quad (24)$$

$$\text{where} \quad \Pi_{z^k + \sigma \Gamma v^k}(\tilde{z}_i) = \begin{cases} \tilde{z}_i & \text{if} \quad |z^k + \sigma \Gamma v^k|_i \leqslant \lambda + \beta, \\ 0 & \text{otherwise.} \end{cases}$$

with $\beta \geqslant 0$. Note that when $\beta = 0$, deriving z^k, u^k and v^k for almost all f at f in the direction f using the chain rule leads to the sequences $\tilde{z}^k$, $\tilde{u}^k$ and $\tilde{v}^k$ respectively (see also [6]). However, as shown in Theorem 2, it is important to choose $\beta > 0$ to guarantee the convergence of the sequence[1].

Theorem 2 *Let $\alpha > 0$ be the minimum non zero value[2] of $|\Gamma u^\star|_i$ for all $i \in [L]$. Choose β such that $\alpha \sigma > \beta > 0$. The sequence $\tilde{u}^k$ defined in (24) converges to the debiasing $\tilde{u}^\star$ of $u^\star$.*

Before turning to the proof of this theorem, let us introduce a first lemma.

Lemma 1 *The debiasing $\tilde{u}^\star$ of $u^\star$ is the solution of the saddle-point problem*

$$\min_{\tilde{u} \in \mathbb{R}^N} \max_{\tilde{z} \in \mathbb{R}^L} \|\Phi \tilde{u} - f\|^2 + \langle \Gamma \tilde{u}, \tilde{z} \rangle - \iota_{F_{\mathcal{I}}}(\tilde{z}), \quad (25)$$

where $\iota_{F_{\mathcal{I}}}$ is the indicator function of the convex set $F_{\mathcal{I}} = \left\{ p \in \mathbb{R}^L \; ; \; p_{\mathcal{I}} = 0 \right\}$.

Proof. As ΦU has full column rank, the debiased solution is the unique solution of the constrained least square estimation problem

$$\tilde{u}^\star = U(\Phi U)^+ f = \operatorname*{argmin}_{\tilde{u} \in \mathcal{U}^\star} \|\Phi \tilde{u} - f\|^2 . \quad (26)$$

Remark that $\tilde{u} \in \mathcal{U}^\star = \mathrm{Ker}[\mathrm{Id}_{\mathcal{I}^c}^t \Gamma] \Leftrightarrow (\Gamma \tilde{u})_{\mathcal{I}^c} = 0 \Leftrightarrow \iota_{F_{\mathcal{I}^c}}(\Gamma \tilde{u}) = 0$, where $F_{\mathcal{I}^c} = \left\{ p \in \mathbb{R}^L \; ; \; p_{\mathcal{I}^c} = 0 \right\}$.

Using Fenchel transform, $\iota_{F_{\mathcal{I}^c}}(\Gamma \tilde{u}) = \max_{\tilde{z}} \langle \Gamma \tilde{u}, \tilde{z} \rangle - \iota_{F_{\mathcal{I}^c}}^*(\tilde{z})$, where $\iota_{F_{\mathcal{I}^c}}^*$ is the convex conjugate of $\iota_{F_{\mathcal{I}^c}}$. Observing that $\iota_{F_{\mathcal{I}}} = \iota_{F_{\mathcal{I}^c}}^*$ concludes the proof. $\square$

Given Lemma 1, replacing $\Pi_{z^k + \sigma \Gamma v^k}$ in (24) by the projection onto $F_{\mathcal{I}}$, i.e.,

$$\Pi_{F_{\mathcal{I}}}(\tilde{z})_{\mathcal{I}^c} = \tilde{z}_{\mathcal{I}^c} \quad \text{and} \quad \Pi_{F_{\mathcal{I}}}(\tilde{z})_{\mathcal{I}} = 0 , \quad (27)$$

leads to the primal-dual algorithm of [4] applied to problem (25) which converges to the debiased estimator $\tilde{u}^\star$. It remains to prove that the projection $\Pi_{z^k + \sigma \Gamma v^k}$ defined in (24) converges to $\Pi_{F_{\mathcal{I}}}$ in finite time.

Proof (Theorem 2). First consider $i \in \mathcal{I}$, i.e., $|\Gamma u^\star|_i > 0$. By assumption on α, $|\Gamma u^\star|_i \geq \alpha > 0$. Necessary $z_i^\star = \lambda \operatorname{sign}(\Gamma u^\star)_i$ in order to maximize (21). Hence, $|z^\star + \sigma \Gamma u^\star|_i \geq \lambda + \sigma \alpha$. Using the triangle inequality shows that

$$\lambda + \sigma \alpha \leqslant |z^\star + \sigma \Gamma u^\star|_i \leqslant |z^\star - z^k|_i + \sigma |\Gamma u^\star - \Gamma v^k|_i + |z^k + \sigma \Gamma v^k|_i . \quad (28)$$

---

[1] In practice, β can be chosen as the smallest positive floating number.
[2] If $|\Gamma u^\star|_i = 0$ for all $i \in [L]$, the result remains true for any $\alpha > 0$.

Choose $\varepsilon > 0$ sufficiently small such that $\sigma\alpha - \varepsilon(1+\sigma) > \beta$. From the convergence of the primal-dual algorithm of [4], the sequence (z^k, u^k, v^k) converges to $(z^\star, u^\star, u^\star)$. Therefore, for k large enough, $|z^\star - z^k|_i < \varepsilon$, $|\Gamma u^\star - \Gamma v^k|_i < \varepsilon$, and

$$|z^k + \sigma\Gamma v^k|_i \geqslant \lambda + \sigma\alpha - \varepsilon(1+\sigma) > \lambda + \beta \ . \tag{29}$$

Next consider $i \in \mathcal{I}^c$, i.e., $|\Gamma u^\star|_i = 0$, where by definition $|z^\star|_i \leqslant \lambda$. Using again the triangle inequality shows that

$$|z^k + \sigma\Gamma v^k|_i \leqslant |z^k - z^\star|_i + \sigma|\Gamma v^k - \Gamma u^\star|_i + |z^\star|_i \ . \tag{30}$$

Choose $\varepsilon > 0$ sufficiently small such that $\varepsilon(1+\sigma) < \beta$. As $(z^k, u^k, v^k) \to (z^\star, u^\star, u^\star)$, for k large enough, $|z^k - z^\star|_i < \varepsilon$, $|\Gamma v^k - \Gamma u^\star|_i < \varepsilon$, and

$$|z^k + \sigma\Gamma v^k|_i < \lambda + \varepsilon(1+\sigma) \leqslant \lambda + \beta \ . \tag{31}$$

It follows that for k sufficiently large $|z^k + \sigma\Gamma v^k|_i \leqslant \lambda + \beta$ if and only if $i \in \mathcal{I}^c$, and hence $\Pi_{z^k + \sigma K v^k}(\tilde{z}) = \Pi_{F_\mathcal{I}}(\tilde{z})$. As a result, all subsequent iterations of (24) will solve (25), and hence from Lemma 1 this concludes the proof of the theorem. $\square$

6   Debiasing other affine estimators

In most cases, $U^\star$ cannot be computed in reasonable memory load and/or time, such that Eq. (12) cannot be used directly. However, the directional derivative, i.e., the application of $J^\star$ to a direction δ, can in general be obtained with an algorithm of the same complexity as the one providing $u^\star$. If one can compute the directional derivatives for any direction, a general iterative algorithm for the computation of $\tilde{u}^\star$ can be derived as given in Algorithm 1.

The proposed technique relies on the fact that given $n = \dim(\mathcal{M}^\star)$ uniformly random directions $\delta_1, \ldots, \delta_n$ on the unit sphere of $\mathbb{R}^P$, $J^\star\delta_1, \ldots, J^\star\delta_n$ forms a basis of $\text{Im}[J^\star]$ almost surely. Given this basis, the debiased solution can so be retrieved from (12). Unfortunately, computing the image of the usually large number n of random directions can be computationally prohibitive.

The idea is to approach the debiased solution by retrieving only a low dimensional subspace of $\mathcal{M}^\star$ leading to a small approximation error. Our greedy heuristic is to chose random perturbations around the current residual (the strength of the perturbation being controlled by a parameter ε). As soon as $\varepsilon > 0$, the algorithm converges in n iterations as explained above. But, by focusing in directions guided by the current residual, the algorithm refines in priority the directions for which the current debiasing gets significantly away from the data f, i.e., directions that encodes potential remaining bias. Hence, the debiasing can be very effective even though a small number of such directions has been explored. We notice in our experiments that with a small value of ε, this strategy leads indeed to a satisfying debiasing, close to convergence, reached in a few iterations.

The nonlocal-means example. The block-wise nonlocal-means proposed in [2] can be rewritten as an instance of the minimization problem (1) with

$$E(u, f) = \frac{1}{2}\sum_{i,j} w_{i,j}\|\mathcal{P}_i u - \mathcal{P}_j f\|^2 \quad \text{with} \quad w_{i,j} = \varphi\left(\frac{\|\mathcal{P}_i f - \mathcal{P}_j f\|^2}{2\sigma^2}\right) \tag{32}$$

Algorithm 1 General debiasing pseudo-algorithm for the computation of $\tilde{u}^\star$.

Inputs: $f \in \mathbb{R}^P$, $u^\star \in \mathbb{R}^N$, $\delta \in \mathbb{R}^P \to J^\star \delta \in \mathbb{R}^N$, $\varepsilon > 0$.
Outputs: $\tilde{u}^\star \in \mathbb{R}^N$ and $U \in \mathbb{R}^{N \times n'}$ with $n' \leqslant n$ an orthonormal family of $\mathrm{Im}[J^\star]$

Initialize $U \leftarrow [\,]$
repeat until $\tilde{u}^\star$ reaches convergence
    Generate $\delta \leftarrow \eta/\|\eta\|, \eta \sim \mathcal{N}_P(0, \mathrm{Id})$ (perturbation ensuring convergence)
    Compute $u' \leftarrow J^\star(f - \Phi\tilde{u}^\star + \varepsilon\delta)$ (perturbed image of the current residual)
    Compute $e \leftarrow u' - U(U^t u')$ (projection on the orthogonal of the current $\mathcal{M}^\star$)
    Update $U \leftarrow [U\ e/\|e\|]$
    Update $\tilde{u}^\star \leftarrow u^\star + U((\Phi U)^+(f - \Phi u^\star))$

---

where $i \in [n_1] \times [n_2]$ spans the whole image domain, $j - i \in [-s, s] \times [-s, s]$ spans a limited search window domain and σ^2 is the noise variance. We denote by $\mathcal{P}_i$ the linear operator extracting the patch at pixel i of size $(2p + 1) \times (2p + 1)$. Note that we assume periodic conditions such that all quantities remain inside the image domain. The kernel $\varphi : \mathbb{R}^+ \to [0, 1]$ is a decreasing function which is typically a decay exponential function. Taking φ piece-wise constant[3], leads to computing $u^\star$ and its Jacobian at f applied to δ for almost all f as follow

$$u_i^\star = \frac{\sum_j \bar{w}_{i,j} f_j}{\sum_j \bar{w}_{i,j}} \quad \text{and} \quad (J^\star \delta)_i = \frac{\sum_j \bar{w}_{i,j} \delta_j}{\sum_j \bar{w}_{i,j}} \quad \text{with} \quad \bar{w}_{i,j} = \sum_k w_{i-k, j-k} \quad (33)$$

where $k \in [-p, p] \times [-p, p]$ spans the patch domain. Note that the values of w and $\bar{w}$ can be obtained by discrete convolutions leading to an algorithm with complexity in $O(Ns^2)$, independent of the half patch size p.

With such a choice of φ, the block-wise nonlocal filter becomes a piece-wise affine mapping of f and hence Algorithm 1 applies.

7   Numerical experiments and results

Figure 1 gives an illustration of TV used for denoising a 1D piece-wise constant signal in $[0, 192]$ and damaged by additive white Gaussian noise (AWGN) with a standard deviation $\sigma = 10$. Even though TV has perfectly retrieved the support of ∇u_0 with one more extra jump, the intensities of some regions are biased. Our debiasing is as expected unbiased for every region.

Figure 2 gives an illustration of our debiasing of 2D anisotropic TV used for the restoration of an $8bits$ approximately piece-wise constant image damaged by AWGN with $\sigma = 20$. The observation operator Φ is a Gaussian convolution kernel of bandwidth 2px. TV introduced a significant loss of contrast, typically for the thin contours of the drawing, which are re-enhanced by our debiased result.

Figure 3 gives an illustration of our iterative debiasing for the block-wise nonlocal-means algorithm used in a denoising problem for an $8bits$ image enjoying many repetitive patterns and damaged by AWGN with $\sigma = 20$. Convergence has been considered as reached after 4 iterations only. Our debiasing provides favorable results with many enhanced details compared to the biased result.

---

[3] For instance, by quantification on a subset of predefined values in $[0, 1]$.
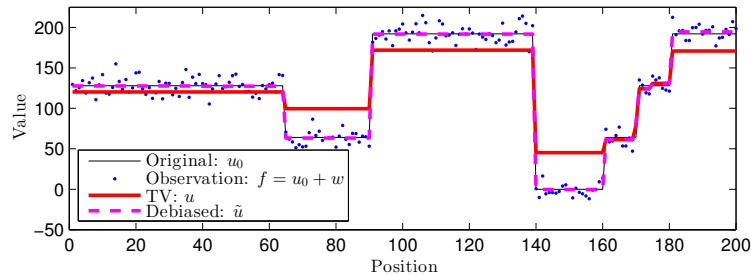
Fig. 1. Solutions of 1D-TV and our debiasing on a piece-wise constant signal.

PSNR 19.13 / SSIM 0.76   PSNR 20.61 / SSIM 0.80   PSNR 21.90 / SSIM 0.87

Fig. 2. (left) Blurry image $f = \Phi u_0 + w$, (center) TV $u^\star$, (right) debiased $\tilde{u}^\star$.

PSNR 22.14 / SSIM 0.52   PSNR 27.89 / SSIM 0.82   PSNR 29.17 / SSIM 0.87

Fig. 3. (left) Noisy image $f = u_0 + w$, (center) nonlocal-means $u^\star$, (right) debiased $\tilde{u}^\star$.

8   Conclusion

We have introduced in this paper a mathematical definition of debiasing which has led to an effective debiasing technique that can remove the method bias that does not arise from the unavoidable choice of the model. This debiasing technique simply consists in applying a least-square estimation constrained to the model subspace chosen implicitly by the original biased algorithm. Numerical experiments have demonstrated the efficiency of our technique in retrieving the correct intensities while respecting the structure of the original model subspace. Our technique is nevertheless limited to locally affine estimators. Isotropic total variation, structured sparsity or nonlocal-means with smooth kernels are not yet handled by our debiasing technique, and left for future work.

References

1. T. Blumensath and M. E. Davies. Iterative thresholding for sparse approximations. J. Fourier Anal. Appl., 14(5-6):629–654, 2008.

2. A. Buades, B. Coll, and J.-M. Morel. A review of image denoising algorithms, with a new one. SIAM J. Multiscale Model. Simul., 4(2):490–530, 2005.
3. M. Burger, G. Gilboa, S. Osher, J. Xu, et al. Nonlinear inverse scale space methods. Communications in Mathematical Sciences, 4(1):179–212, 2006.
4. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis., 40:120–145, 2011.
5. S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. SIAM J. Sci. Comput., 20(1):33–61, 1998.
6. C.-A. Deledalle, S. Vaiter, G. Peyré, and J. M. Fadili. Stein unbiased gradient estimator of the risk (SUGAR) for multiple parameter selection. SIAM J. Imaging Sciences, 7(4):2448–2487, 2014.
7. B. Denis De Senneville, S. Roujol, S. Hey, C. Moonen, and M. Ries. Extended Kalman filtering for continuous volumetric MR-temperature imaging. IEEE Trans. Med. Imaging, 32(4):711–718, 2013.
8. D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. Biometrika, 81(3):425–455, 1994.
9. B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. Ann. Statist., 32(2):407–499, 2004.
10. M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. Inverse problems, 23(3):947, 2007.
11. J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Am. Statist. Assoc., 96(456):1348–1360, 2001.
12. G. Gilboa. A total variation spectral framework for scale and texture analysis. SIAM J. Imaging Sciences, 7(4):1937–1961, 2014.
13. K. K. Herrity, A. C. Gilbert, and J. A. Tropp. Sparse approximation via iterative thresholding. In ICASSP, volume 3, pages III–III. IEEE, 2006.
14. A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970.
15. J. Lederer. Trust, but verify: benefits and pitfalls of least-squares refitting in high dimensions. arXiv preprint arXiv:1306.0113, 2013.
16. C. Louchet and L. Moisan. Total variation as a local filter. SIAM Journal on Imaging Sciences, 4(2):651–694, 2011.
17. S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. SIAM J. Multiscale Model. Simul., 4(2):460–489, 2005.
18. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena, 60(1):259–268, 1992.
19. D. Strong and T. Chan. Edge-preserving and scale-dependent properties of total variation regularization. Inverse problems, 19(6):S165, 2003.
20. R. Tibshirani. Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B, pages 267–288, 1996.
21. A. N. Tikhonov. On the stability of inverse problems. Dokl. Akad. Nauk SSSR, 39:176–179, 1943.
22. S. Vaiter, C.-A. Deledalle, G. Peyré, C. Dossal, and J. Fadili. Local behavior of sparse analysis regularization: Applications to risk estimation. Appl. Comput. Harmon. Anal., 35(3):433–451, 2013.
23. S. Vaiter, M. Golbabaee, M. Fadili, and G. Peyré. Model selection with low complexity priors. arXiv preprint arXiv:1307.2342, 2014.
24. J. Xu and S. Osher. Iterative regularization and nonlinear inverse scale space applied to wavelet-based denoising. IEEE Trans. Image Proc., 16(2):534–544, 2007.