Super-Resolution with Sparse Mixing Estimators

Stéphane MALLAT and Guoshen YU

CMAP, Ecole Polytechnique, 91128 Palaiseau Cedex, France

October 23, 2009

Abstract

We introduce a class of inverse problem estimators computed by mixing adaptively a family of linear estimators corresponding to different priors. Sparse mixing weights are calculated over blocks of coefficients in a frame providing a sparse signal representation. They minimize an $\mathbf{l}^1$ norm taking into account the signal regularity in each block. Adaptive directional image interpolations are computed over a wavelet frame with an $O(N \log N)$ algorithm.

I. INTRODUCTION

Many signal acquisition and restoration require to solve an inverse problem while trying to improve the signal resolution. It amounts to estimate a high resolution signal $f \in \mathbf{R}^N$ from Q measurements $y[q]$, obtained through a linear operator U, and contaminated by an additive noise w

$$y[q] = Uf[q] + w[q] \quad q \in \mathcal{G} \quad \text{with} \quad |\mathcal{G}| = Q < N .$$

Image interpolation is an important example, where U is a subsampling operator. Many image display devices have zooming abilities that interpolate input images to adapt their size to high resolution screens. For example, high definition televisions include a spatial interpolator which increases the size of standard definition videos to match the high definition screen format and possibly improve the image quality.

Linear operators compute an estimator $\tilde{f}$ which also belongs to a space of dimension Q, and thus does not improve the signal resolution. For image interpolations, bicubic interpolators most often provide nearly the best results among linear operators [35]. To estimate f in a space of dimension larger than Q requires using non-linear estimators adapted to prior information on the signal properties. A wide range of techniques have been developed to improve linear image interpolators. Directional image interpolations take advantage of the geometric regularity of image structures by performing the interpolation in a chosen direction along which the image

is locally regular. The main difficulty is to locally identify this direction of regularity. Along an edge, the interpolation direction should be parallel to the edge. Many adaptive interpolations have been thus developed with edge detectors [33], [2], [51] and by finding the direction of regularity with gradient operators [62], [36], [13], [59], [34], [18], [7], [1], [6], [16], [58]. More global image models impose image smoothness priors such as a bounded total variation to optimize the interpolation [39], [44], [3], [43]. Other image smoothness priors have also been used to compute interpolated images with alternate projections on convex sets [49], [9]. These algorithms can provide a better image quality then a linear interpolator but they also produces artifacts so that the resulting PSNR remains of the same order as a bicubic interpolator. The introduction of interpolators adapted to local covariance image models have lead to more precise estimators [38]. This approach has been improved by Zhang and Wu [63] by using autoregressive image models optimized over image blocks. In most cases, it currently provides the best PSNR for spatial image interpolation. Super-resolution interpolations can further be improved by using a sequence of images [48], [30], [31], [46], [27] or a comparison dataset [28], [22], [58] to perform the interpolation. While these approaches can be more accurate, they are much more demanding in computation and memory resources.

Prior information on the image sparsity has also been used for image interpolation. Wavelet estimators were introduced to compute fine scale wavelet coefficients by extrapolating larger scale wavelet coefficients [12], [15]. A more general and promising class of non-parametric super-resolution estimators assumes that the high resolution signal f is sparse in some dictionary of vectors. This sparse representation is estimated by decomposing the low-resolution measurements y in a transformed dictionary. These algorithms, which are reviewed in Section II, have found important applications for sparse spike inversion in geophysics or image inpainting [23], [26]. However, they do not provide state-of-the-art results for image interpolation.

Section III describes a new class of adaptive inverse estimators, calculated over a sparse signal representation in a frame. It is obtained with a sparse adaptive mixing of a family of linear estimators, which are optimized for different signal priors. Mixing linear estimators has been shown to be very effective for noise removal [37]. However, these approaches do not apply to inverse problems because they rely on a Stein unbiased empirical estimator of the risk, which is then not valid.

Our inverse sparse mixing estimator is derived from a mixture model of the measurements y.

It is computed in Section IV by minimizing an $\mathbf{l}^1$ norm over blocks of frame coefficients, with weights depending upon the different signal priors. Section V describes a fast block orthogonal matching pursuit algorithm which computes the mixing weights. Linear mixture models have been studied over wavelet coefficients for image denoising [47]. For image interpolation, Section VI implements the inverse mixing estimator in a wavelet frame with $O(N \log N)$ operations, with state-of-the-art numerical results.

II. Sparse Inverse Problem Estimation in Dictionaries

Sparse super-resolution estimations over dictionaries provide effective non parametric approaches to inverse problems. These algorithms are reviewed with their application to image interpolation.

A signal $f \in \mathbf{R}^N$ is estimated by taking advantage of prior information which specifies a dictionary $\mathscr{D} = \{\phi_p\}_{p \in \Gamma}$ where f has a sparse approximation. This dictionary may be a basis or some redundant frame, with a size $|\Gamma| = P \geq N$. Sparsity means that f is well approximated by its orthogonal projection f_Λ over a sub-space $\mathbf{V}_\Lambda$ generated by a small number $M = |\Lambda|$ of vectors $\{\phi_p\}_{p \in \Lambda}$ chosen in $\mathscr{D}$:

$$f_\Lambda = \sum_{p \in \Lambda} c(p)\, \phi_p. \tag{1}$$

Measurements are obtained with a linear operator U, with an additive noise w:

$$y(q) = Uf(q) + w(q) \quad \text{for} \quad q \in \mathscr{G}, \quad \text{with} \quad |\mathscr{G}| = Q < N . \tag{2}$$

Sparse inversion algorithms estimate the approximation space $\mathbf{V}_\Lambda$ of f from y, together with the decomposition coefficients $c(p)$ of the projection of f in $\mathbf{V}_\Lambda$. It results from (1) and (2) that

$$y = \sum_{p \in \Lambda} c(p) U \phi_p + w' \quad \text{with} \quad w' = U(f - f_\Lambda) + w . \tag{3}$$

This means that y is well approximated by a projection in a space $U \mathbf{V}_\Lambda = \{U \phi_p\}_{p \in \Lambda}$. The space $\mathbf{V}_\Lambda$ and the coefficients $c(p)$ are estimated by finding a sparse representation of y in the transformed dictionary

$$\mathscr{D}_U = \{U \phi_p\}_{p \in \Gamma} . \tag{4}$$

All vectors $U \phi_p$ belong to the image space of U, which is of dimension Q. Since there are $P \geq N > Q$ such vectors, the transformed dictionary $\mathscr{D}_U$ is redundant, and y has an infinite

number of possible decompositions in this dictionary. A sparse approximation

$$\tilde{y} = \sum_{p \in \tilde{\Lambda}} \tilde{c}(p) U \phi_p \ . \tag{5}$$

can be calculated with a basis pursuit algorithm which minimizes a Lagrangian penalized by an $\mathbf{l}^1$ norm [53], [14]

$$\frac{1}{2} \|y - \sum_{p \in \Gamma} \tilde{c}(p) U g_p\|^2 + \lambda \sum_{p \in \Gamma} |\tilde{c}(p)| \ . \tag{6}$$

A sparse representation can also be calculated with faster greedy matching pursuit algorithms [42]. Let $\tilde{\Lambda}$ be the support of $\tilde{c}(p)$ in Γ. The resulting sparse estimation $\tilde{f}$ of f is given by

$$\tilde{f} = \sum_{p \in \tilde{\Lambda}} \tilde{c}(p) \phi_p \ . \tag{7}$$

Such an estimation is precise and stable if the support $\tilde{\Lambda}$ of $\tilde{c}$ includes a precise approximation support Λ of the decomposition coefficients of f, so that it recovers an estimator $\tilde{f}$ in a space $\mathbf{V}_{\tilde{\Lambda}} \subset \mathbf{V}_\Lambda$. One must also guarantee that the computations are stable and hence that $\{U\phi_p\}_{p \in \Lambda}$ is a Riesz basis. The "Restrictive Isometry Property" of Candes and Tao [11] and Donoho [21] imposes that the Riesz constants are uniformly bounded for all supports of a given size. They then proved that the recovery is precise and stable. This restrictive isometry property is valid for certain classes of random operators U but not for structured operators such as a subsampling on a uniform grid. For structured operators, the precision and stability of this sparse inverse estimation depends upon the "geometry" of Λ, which is not well understood mathematically, despite some sufficient exact recovery conditions proved by Tropp [55], [56].

Several authors have applied this sparse super-resolution algorithm for image interpolation and inpainting. Curvelet frames [10] and contourlet frames [20] build sparse image approximations by taking advantage of the image directional regularity. Dictionaries of curvelet frames have been applied successfully to image inpainting [23], [26]. For uniform grid interpolations, Table I in Section VI shows that the resulting estimations are not as precise as linear bicubic interpolations. Table I shows that a contourlet algorithm [45] sometimes can provide a slightly better PSNR then a bicubic interpolation, but these results are below the state of the art obtained with adaptive directional interpolators [63]. Dictionaries of image patches have also been studied for image interpolations with sparse representations [60], but with little PSNR improvements compared to bicubic interpolations.

A source of instability of these algorithms come from their flexibility, which does not incorporate enough prior information. The approximation space $\mathbf{V}_\Lambda$ is estimated by selecting independently each of the dictionary vector ϕ_p. A selection of M vectors thus corresponds to a choice of an approximation space among $\binom{P}{M}$ possible subspaces. It does not take into account geometric image structures which create dependencies on the choice of approximation vectors. Structured approximation algorithms use such prior information to restrict the set of possible approximation spaces [29], [32]. Since approximation vectors often appear in groups, one can select simultaneously blocks of approximation vectors [25], [24], [52], which reduces the number of possible approximation spaces. The $\mathbf{l}^1$ penalization in (6) is then replaced by a sum of the $\mathbf{l}^2$ norm over each block, which results in a mixed $\mathbf{l}^1$ and $\mathbf{l}^2$ norm [65]. This is also called a "group lasso" optimization [61], [50], [4]. These structured approximations have been shown to improve the signal estimation in a compressive sensing context for a random operator U [5], [24]. However, for more unstable inverse problems such as image interpolation, this regularization is not sufficient to reach state-of-the-art results.

III. MIXING ESTIMATORS OVER FRAME BLOCKS

Sparse super-resolution algorithms can be improved by using more prior information on the signal properties. This section introduces a general class of sparse inverse estimators that define signal approximations over blocks of vectors in a frame. This class of estimators are introduced as an adaptive mixing of linear estimators.

A Tikhonov regularization optimizes a linear estimator by imposing that the solution has a regularity specified by a quadratic prior [54]. Suppose that f has a regularity which is measured by a quadratic regularity norm $\|R_\theta f\|^2$, where R_θ is a linear operator. Sobolev norms are particular examples where R_θ are differential operators. Let σ^2 be the variance of the noise w. A Tikhonov estimator computes $\tilde{f} = U_\theta^+ y$ by minimizing $\|R_\theta \tilde{f}\|^2$ subject to

$$\|U\tilde{f} - y\|^2 \leq Q\sigma^2 \ . \tag{8}$$

The solution of this quadratic minimization problem is also obtained by minimizing a Lagrangian

$$\frac{1}{2}\|U\tilde{f} - y\|^2 + \lambda \|R_\theta \tilde{f}\|^2 \ . \tag{9}$$

In Bayesian terms, this Lagrangian is minus the log of the posterior distribution of the signal given the observations y, whose minimization yields a maximum a posterior estimator. The first

term is proportional to minus the log probability of the Gaussian distribution of the noise. The second term is minus the log probability of a Gaussian prior distribution whose covariance is $(R_\theta^* R_\theta)^{-1}$, where R_θ^* is the adjoint of R_θ. In this framework, the regularity prior is thus interpreted as a covariance prior.

If we neglect the noise, which is often the case for image interpolation then (8) becomes $UU_\theta^+ y = y$. If U is a subsampling operator by a factor 2 and if R_θ is a convolution operator then the resulting Tikhonov interpolator U_θ^+ implements a linear filtering with an impulse response h_θ:

$$\tilde{f}(n) = U_\theta^+ y(n) = \sum_{q \in \mathscr{G}} y(q) \, h_\theta(n - 2q) \quad \text{for} \quad n = (n_1, n_2).$$

Let $\mathbf{V}_\theta$ be the image space of a linear estimator U_θ^+. The estimation $\tilde{f} = U_\theta^+ y$ can only be a precise approximation if f is well approximated by its projection in $\mathbf{V}_\theta$. Adaptive estimators introduce more flexibility on the construction of this approximation space, which is obtained as a union of subspaces selected depending upon the signal regularity. We introduce a class of such estimators by estimating a mixture model of f.

A global linear mixture would decompose f as a combination of signals having a regularity R_θ. This is equivalent to model f as a realization of a linear mixture of Gaussian random vectors with covariances $(R_\theta^* R_\theta)^{-1}$. The regularity $\|R_\theta f\|^2$ is estimated from y by minimizing $\|R_\theta \tilde{f}\|^2 = \|\tilde{R}_\theta y\|^2$ which is obtained with the Tikhonov estimator U_θ^+:

$$\tilde{R}_\theta = R_\theta U_\theta^+ \ .$$

To adapt the mixture to local signal properties, it is not computed globally but locally in a basis or frame providing a sparse signal representation. A sparse representation reduces the estimation to lower dimensional spaces where the signal projection is non-negligible.

Let us consider a basis or frame $\{\psi_p\}_{p \in \Gamma}$ and its dual frame $\{\tilde{\psi}_p\}_{p \in \Gamma}$, which provide a sparse representation of y

$$y = \sum_{p \in \Gamma} c(p) \, \tilde{\psi}_p \quad \text{with} \quad c(p) = \langle f, \psi_p \rangle \ .$$

Suppose that we are given a family of regularization operators $\{R_\theta\}_{\theta \in \Theta}$ specifying different signal priors. We define a mixture of y with components having different regularities $\tilde{R}_\theta = R_\theta U_\theta^+$. For each θ, we consider index blocks $B_{\theta,q} \subset \Gamma$ where q is a position parameter that is sampled

over a subgrid Γ_θ of Γ. These blocks cover the index set Γ but may have a non-empty intersection

$$\cup_{q\in\Gamma_\theta} B_{\theta,q} = \Gamma \ .$$

Let $\mathbf{V}_{\theta,q}$ be the space generated by $\{\psi_p\}_{p\in B_{\theta,q}}$. We want to select blocks $B_{\theta,q}$ where the signal projection

$$y_{\theta,q} = \sum_{p\in B_{\theta,q}} c(p)\,\tilde{\psi}_p \in \mathbf{V}_{\theta,q} \tag{10}$$

has mostly a regularity $\tilde{R}_\theta$ or can be interpreted as the realization of a Gaussian vector whose covariance is dominated by $(\tilde{R}_\theta^* \tilde{R}_\theta)^{-1}$. It defines an adaptive local signal mixture over blocks

$$y = \sum_{\theta\in\Theta}\sum_{q\in\Gamma_\theta} \tilde{a}(\theta,q)\,y_{\theta,q} + y_r \ . \tag{11}$$

Each block $B_{\theta,q}$ is selected if the mixing coefficient $\tilde{a}(\theta,q)$ is close to 1 and it is removed if $\tilde{a}(\theta,q)$ is close to 0. The residual signal y_r is not dominated by one component and thus has no specific regularity.

Let $\{U_\theta^+\}_{\theta\in\Theta}$ be the optimal linear Tikhonov estimators corresponding to the priors $\{R_\theta\}_{\theta\in\Theta}$. For interpolation, the U_θ^+ are interpolators in several directions θ. A mixture estimator is defined from a mixture model (11) by inverting each signal component of prior R_θ with U_θ^+ and the residue with a generic estimator U^+:

$$\tilde{f} = \sum_{\theta\in\Theta} U_\theta^+ \left(\sum_{q\in\Gamma_\theta} \tilde{a}(\theta,q)\,y_{\theta,q} \right) + U^+(y_r) \ . \tag{12}$$

The generic linear estimator U^+ does not incorporate any prior knowledge concerning the R_θ signal regularity. In a Bayesian framework, U^+ is an estimator computed with a prior Gaussian distribution whose covariance is not conditioned on θ. It can be computed from the covariances $(R_\theta^* R_\theta)^{-1}$ of each prior distribution conditioned upon θ, and from the probability distribution of θ. For image interpolations, U^+ is isotropic, or nearly isotropic if implemented with a separable interpolation such as a bicubic interpolation.

Inserting (10) in (12) yields a mixing estimator which locally adapts the inverse operator to the signal regularity:

$$\tilde{f} = \sum_{p\in\Gamma} c(p) \left(\sum_{\theta\in\Theta} a_\theta(p)\,U_\theta^+ + a_r(p)\,U^+ \right) \tilde{\psi}_p, \tag{13}$$

with mixing weights

$$a_\theta(p) = \sum_{q\in\Gamma_\theta} \tilde{a}(\theta,q)\,1_{B_{\theta,q}}(p) \quad\text{and}\quad a_r(p) = 1 - \sum_{\theta\in\Theta} a_\theta(p) \ . \tag{14}$$

IV. Sparse Mixture Estimation

The choice of a mixing estimator (12) is derived from a mixture model of y:

$$y = \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta, q) y_{\theta, q} + y_r \quad \text{with} \quad y_{\theta, q} = \sum_{p \in B_{\theta, q}} c(p) \tilde{\psi}_p . \tag{15}$$

Computing such a model can be interpreted as a non-standard source separation problem, with only one measurement channel. The mixing parameters $\tilde{a}(\theta, q)$ must be estimated from a known set of potential sources $y_{\theta, q}$ which are highly redundant, with a prior information on their quadratic regularity. A sparse mixing estimator is introduced with a weighted $\mathbf{l}^1$ norm optimization.

A linear mixture estimator can be obtained by minimizing the residue energy $\|y_r\|^2$ penalized by the signal regularity over all blocks measured by

$$\sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \|\tilde{R}_\theta \tilde{a}_{\theta, q} y_{\theta, q}\|^2 = \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} |\tilde{a}_{\theta, q}|^2 \|\tilde{R}_\theta y_{\theta, q}\|^2. \tag{16}$$

However, this approach does not take advantage of the sparsity prior.

Since the signal has a sparse representation, many block signals $y_{\theta, q}$ are close to zero. If it is not the case then the signal model assumes that they have a regularity specified by one of the operators R_θ. It implies that the mixing coefficients $\tilde{a}(\theta, q)$ should locally be non-negligible for one or no parameter θ, and is therefore sparse. Sparsity priors have been used in standard blind source separation problems [64], [8], with a sparsity prior on the unknown sources. In this case the sparsity is not imposed on the sources but on mixing coefficients. According to the sparsity approach reviewed in Section II, the quadratic prior norm on mixing coefficients in (16) is thus replaced by an $\mathbf{l}^1$ norm. Mixing coefficients are obtained by minimizing the residual norm $\|y_r\|^2$ penalized by the resulting weighted $\mathbf{l}^1$ prior

$$\mathcal{L}(\tilde{a}) = \frac{1}{2} \|y - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta, q) y_{\theta, q}\|^2 + \lambda \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} |\tilde{a}(\theta, q)| \|\tilde{R}_\theta y_{\theta, q}\|^2. \tag{17}$$

The minimization of such a quadratic function of the unknown $\tilde{a}(\theta, q)$ penalized by their $\mathbf{l}^1$ norm can be computed with standard algorithms, such as an iterative thresholding [19].

As opposed to group lasso algorithms using mixed $\mathbf{l}^2$ and $\mathbf{l}^1$ norms, this minimization does not only recover the signal with a sparse set of blocks but it also regularizes the decomposition by imposing a signal regularity within each block. Moreover, it does not optimize a decomposition

parameter for each frame coefficient but a single mixing parameter per block. The signal regularity in each block can also be interpreted as a linear approximation property in an orthonormal basis defined in the block. Let us denote

$$\tilde{R}_{\theta,q} y = \tilde{R}_\theta y_{\theta,q} = \sum_{p \in B_{\theta,q}} c(p) \tilde{R}_\theta \tilde{\psi}_p .$$

It results that

$$\|\tilde{R}_\theta y_{\theta,q}\|^2 = \langle \tilde{R}^*_{\theta,q} \tilde{R}_{\theta,q} y, y \rangle .$$

The symmetric operator $\tilde{R}^*_{\theta,q} \tilde{R}_{\theta,q}$ is diagonalized in $\mathbf{V}_{\theta,q}$ in an orthonormal basis $\{b_{\theta,q,m}\}_m$ with eigenvalues λ_m^2:

$$\tilde{R}^*_{\theta,q} \tilde{R}_{\theta,q} y = \sum_m \lambda_m^2 \langle y, b_{\theta,q,m} \rangle b_{\theta,q,m}.$$

If the eigenvalues λ_m^2 vary by a large factor then the energy $\|\tilde{R}_\theta y_{\theta,q}\|^2 / \|y_{\theta,q}\|^2$ is small if and only if y has an energy concentrated over the eigenvectors $\{b_{\theta,q,m}\}_m$ corresponding to the smallest eigenvalues λ_m^2. The regularity condition is therefore equivalent to a sparse linear approximation condition in this eigenvectors basis.

The blocks $B_{\theta,q}$ have a regularization role in the adaptive selection of estimators U_θ^+ but should not be too large to maintain enough flexibility in the choice of θ. The regularization is effective if the eigenvalues $\{\lambda_m^2\}_m$ vary by a sufficiently large factor so that one can indeed "observe" the signal regularity in each block. The block shape must therefore be adapted accordingly to the properties of R_θ. For directional interpolation in the direction of θ, a better regularization is obtained with blocks elongated in the direction of θ.

The estimation depends upon the grids Γ_θ of the position indexes q of the blocks $B_{\theta,q}$. To reduce this grid effect, the estimation can be computed with several sets of translated grids. Each grid Γ_θ is translated by several vectors $\{\tau_{\theta,i}\}_{1 \leq i \leq I}$: $\Gamma_{\theta,i} = \Gamma_\theta + \tau_{\theta,i}$. For each i, mixing coefficients $\tilde{a}_i(\theta, q)$ are computed with blocks $B_{\theta,q}$ translated on the grid $\Gamma_{\theta,i}$. The final estimator is obtained by averaging these mixing coefficients

$$\tilde{a}(\theta, q) = \frac{1}{I} \sum_{i=1}^{I} \tilde{a}_i(\theta, q) . \tag{18}$$

V. Computations and Orthogonal Block Matching Pursuit

To reduce the computation of mixing coefficients, an upper bound of the Lagrangian (17) is computed from the frame coefficients of f in each block. Efficient algorithms have been developed for $\mathbf{l}^1$ minimization but they remain slow for image processing applications. An orthogonal block matching pursuit algorithm is introduced to approximate the optimization, with much less computations.

The Euclidean norm of coefficients in a block is written:

$$\|c\|^2_{B_{\theta,q}} = \sum_{p \in B_{\theta,q}} |c(p)|^2 \ .$$

Proposition 1. *If B is the upper frame constant of $\{\psi_p\}_{p \in \Gamma}$ then for all y and $\tilde{a}$:*

$$\mathscr{L}(\tilde{a}) \leq B \mathscr{L}_1(\tilde{a}) \tag{19}$$

where

$$
\begin{aligned}
\mathscr{L}_1(\tilde{a}) \ = \ & \frac{1}{2} \sum_{p \in \mathscr{G}} |c(p)|^2 \Big(1 - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) \, 1_{B_{\theta,q}}(p)\Big)^2 \\
& + \frac{\lambda}{B} \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} |\tilde{a}(\theta,q)| \, \|\bar{R}_{\theta,q} c\|^2_{B_{\theta,q}} \ ,
\end{aligned}
\tag{20}
$$

and $\bar{R}_{\theta,q}$ satisfies

$$\forall (p,p') \in B_{\theta,q} \ , \quad \sum_{m \in B_{q,\theta}} \bar{R}_{\theta,q}(p,m) \bar{R}_{\theta,q}(m,p') = \langle \tilde{R}_\theta \tilde{\psi}_p, \tilde{R}_\theta \tilde{\psi}_{p'} \rangle \tag{21}$$

If the frame is an orthonormal basis then $B = 1$ and $\mathscr{L}(\tilde{a}) = \mathscr{L}_1(\tilde{a})$.

Proof: Observe that

$$\|y - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) y_{\theta,q}\|^2 = \| \sum_{p \in \Gamma} c(p)(1 - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) \, 1_{B_{\theta,q}}(p)) \tilde{\psi}_p\|^2.$$

Since B is the upper frame bound of $\{\psi_p\}_{p \in \Gamma}$

$$\|y - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) y_{\theta,q}\|^2 \leq B \sum_{p \in \Gamma} |c(p)|^2 |1 - \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) \, 1_{B_{\theta,q}}(p)|^2 \ . \tag{22}$$

The regularity norm can be written

$$\|\tilde{R}_\theta y_{\theta,q}\|^2 = \sum_{(p,p') \in B^2_{\theta,q}} c(p) \, c(p') \, \langle \tilde{R}_\theta \tilde{\psi}_p, \tilde{R}_\theta \tilde{\psi}_{p'} \rangle$$

The matrix $\{\langle \tilde{R}_\theta \tilde{\psi}_p, \tilde{R}_\theta \tilde{\psi}_{p'} \rangle\}_{(p,p') \in B_{\theta,q}}$ is symmetric positive and can thus be (non uniquely) factorized into $\bar{R}_{\theta,q}^* \bar{R}_{\theta,q}$ where $\bar{R}_{\theta,q}$ satisfies (21). One can thus rewrite

$$\|\tilde{R}_\theta y_{\theta,q}\|^2 = \langle \bar{R}_{\theta,q}^* \bar{R}_{\theta,q} c, c \rangle = \|\bar{R}_{\theta,q} c\|_{B_{\theta,q}}^2.$$

Inserting this together with (22) in (17) proves (19). If the frame is an orthonormal basis then the inequality (22) is an equality with $B = 1$, so $\mathscr{L}(\tilde{a}) = \mathscr{L}_1(\tilde{a})$. $\square$

In the following, mixing coefficients are computed by minimizing the upper bound $\mathscr{L}_1(\tilde{a})$, which is faster to compute from the frame coefficients of f. With the change of variable

$$\bar{a}(\theta, q) = \tilde{a}(\theta, q) \|\bar{R}_{\theta,q} c\|_{B_{\theta,q}}^2$$

the Lagrangian $\mathscr{L}_1(\tilde{a})$ can be rewritten in a standard form

$$\mathscr{L}_1(\tilde{a}) = \frac{1}{2} \|x - \Phi \bar{a}\|^2 + \lambda \|\bar{a}\|_1 = \sum_{p \in \mathscr{G}} |x(p) - \Phi \bar{a}(p)|^2 + \lambda \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} |\bar{a}(\theta, q)| \qquad (23)$$

with $x(p) = |c(p)|$ and

$$\Phi \bar{a}(p) = \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \bar{a}(\theta, q) \frac{|c(p)| \mathbf{1}_{B_{\theta,q}}(p)}{\|\bar{R}_{\theta,q} c\|_{B_{\theta,q}}^2} . \qquad (24)$$

The minimization of (23) is implemented with iterative algorithms such as [19], which all have a computational complexity dominated by the calculation of Φ^* and Φ at each iteration. Let K be the total number of blocks $\{B_{\theta,q}\}_{\theta \in \Theta, q \in \Gamma_q}$ and S be the maximum size of these blocks. We verify from (24) that the operators Φ and Φ^* are computed with $O(KS)$ operations so L iterations of an $\mathbf{l}^1$ minimizer is implemented with $O(KSL)$ operations.

To further reduce the number of operations, a solution is computed with a greedy minimization implementing an orthogonal block matching pursuit. The algorithm is initialized by setting $\tilde{a}(\theta, q) = 0$ and it computes progressively non-zero mixing coefficients $\tilde{a}(\theta, q)$ to minimize $\mathscr{L}_1(\tilde{a})$ at each step.

If a single $\tilde{a}(\theta, q)$ is chosen to be non-zero, then (20) becomes

$$\mathscr{L}_1(\tilde{a}) = \frac{1}{2} \sum_{p \in \mathscr{G}} |c(p)|^2 + \frac{1}{2} \sum_{p \in \mathscr{B}_{\theta,q}} |c(p)|^2 \left(\tilde{a}(\theta, q)^2 - 2\tilde{a}(\theta, q) \right) + \lambda |\tilde{a}(\theta, q)| \|\bar{R}_{\theta,q} c\|_{B_{\theta,q}}^2 .$$

The minimum is thus obtained with a soft thresholding

$$\tilde{a}(\theta, q) = \rho(\theta, q) = \max \left(1 - \lambda \frac{\|\bar{R}_{\theta,q} c\|_{B_{\theta,q}}^2}{\|c\|_{B_{\theta,q}}^2}, 0 \right) . \qquad (25)$$

The corresponding minimum Lagrangian value is

$$\mathcal{L}_1(\tilde{a}) = \frac{1}{2}(\|c\|^2 - e(\theta, q))$$

with

$$e(\theta, q) = \|c\|^2_{B_{\theta,q}} \rho(\theta, q)^2 \ . \tag{26}$$

The minimization of $\mathcal{L}_1(\tilde{a})$ with a single non-zero mixing coefficient is thus obtained by choosing the block index (θ, q) which maximizes $e(\theta, q)$.

An orthogonal matching pursuit algorithm selects one by one blocks that do not intersect. If $\tilde{a}$ has L non-zero coefficients $\{\tilde{a}(\theta_l, q_l)\}_{1 \leq l \leq L}$ corresponding to non-intersecting blocks B_{θ_l, q_l} then we verify similarly that the mixing coefficients which minimize $\mathcal{L}_1(\tilde{a})$ are

$$\tilde{a}(\theta_l, q_l) = \rho(\theta_l, q_l)$$

and

$$\mathcal{L}_1(\tilde{a}) = \frac{1}{2}\left(\|c\|^2 - \sum_{l=1}^{L} e(\theta_l, q_l)\right)$$

with

$$e(\theta_l, q_l) = \|c\|^2_{B_{\theta_l, q_l}} \rho(\theta_l, q_l)^2 \ .$$

At each iteration, to minimize $\mathcal{L}_1(\tilde{a})$, an orthogonal block matching pursuit finds B_{θ_l, q_l} which maximizes $e(\theta, q)$ among all blocks that do not intersect with the previously selected blocks. The resulting algorithm is described below.

1) Initialization: set $l = 0$ and compute

$$\forall \theta \in \Theta, \ \forall q \in \Gamma_\theta \ , \ \ e(\theta, q) = \|c\|^2_{B_{\theta,q}} \rho(\theta, q)^2 \ , \ \ \tilde{a}(\theta, q) = 0 \ . \tag{27}$$

2) Maxima finding:

$$(\theta_l, q_l) = \arg\max_{\theta, q} e(\theta, q) \ \ \text{and set} \ \ \tilde{a}(\theta_l, q_l) = \rho(\theta_l, q_l).$$

3) Energy update: if $e(\theta_l, q_l) > T$ then eliminate all blocks that intersect with B_{θ_l, q_l}

$$\forall \theta \in \Theta, \ \forall q \in \Gamma_\theta \ , \ \ \text{if} \ \ \sum_p 1_{B_{\theta,q}}(p) \, 1_{B_{\theta_l,q_l}}(p) \neq 0 \ \ \text{set} \ \ e(\theta, q) = 0 \ ,$$

set $l = l + 1$ and go back to step 2).

Otherwise stop.

This algorithm computes mixing coefficients $\tilde{a}(\theta, q)$ for all $\theta \in \Theta$ and $q \in \Gamma_\theta$. It stops when there is no sufficiently energetic block compared to a precision threshold T that is typically proportional to the noise variance. In the image interpolation numerical experiments $T = 0$ as the noise is neglected. The following proposition gives an upper bound of the computations and memory requirements required for an efficient implementation of this algorithm, which is described in the proof. The operators $\bar{R}_{\theta, q}$ are said to be sparse if $\bar{R}_{\theta, q} c$ is computed over $B_{\theta, q}$ with $O(|B_{\theta, q}|)$ operations.

Proposition 2. *Over a family of K blocks of maximum size S, an orthogonal block matching pursuit can be implemented in $O(K(S^2 + \log_2 K))$ operations with $O(KS)$ memory. For sparse regularity operators, the computational complexity is $O(K(S + \log_2 K))$.*

Proof: The geometry of all blocks is stored in an array $B(\theta, q)$ which gives the list of all $p \in B_{\theta, q}$. Each list has less then S elements so the array requires $O(KS)$ memory.

Let us build an array $L(p)$ for $p \in \Gamma$ which stores the list of all (θ, q) for which $p \in B_{\theta, q}$. If $|L(p)|$ is the size of such a list then

$$\sum_{p \in \Gamma} |L(p)| = \sum_{p \in \Gamma} \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} 1_{B_{\theta, q}}(p) = \sum_{\theta \in \Theta} \sum_{q \in \Gamma_\theta} \sum_{p \in \Gamma} 1_{B_{\theta, q}}(p) \ .$$

Since $\sum_{p \in \Gamma} 1_{B_{\theta, q}}(p) \le S$ it results that

$$\sum_{p \in \Gamma} |L(p)| \le KS \ .$$

This array thus requires $O(KS)$ memory.

Energy values are stored in an array of size K indexed by (θ, q) and in an ordered heap of size K. A heap is a binary tree data structure which stores the elements of a set and allows to find the maximum element with $O(1)$ operations [17]. The construction of a heap for a set of K elements requires $O(K)$ operations. The array and the heap require $O(K)$ memory, so the total memory is $O(KS)$. If the blocks are highly structured, which is often the case in applications, then we do not need to store $B(\theta, q)$ and $L(p)$ because these lists can be computed analytically from the blocks shape and position parameters (θ, q). The required memory is then only $O(K)$.

At the initialization, the computation of each of the K energy values $e(\theta, q)$ is dominated by the calculation of $\|\bar{R}_{\theta, q} c\|_{B_{\theta, q}}$ which requires at most $O(S^2)$ operations. The total is thus $O(KS^2)$ operations. If $R_{\theta, q}$ is a sparse regularity operator such as a derivative operator, then it is computed with $O(S)$ operations and the total is therefore $O(KS)$.

The selection of the index (θ_l, q_l) corresponding to the largest energy is implemented in $O(1)$ operation by finding the element at the top of the heap. Since there are most K iterations, these steps are implemented in $O(K)$ operations.

For the energy update, for all $p \in B_{\theta_l, q_l}$ stored in $B(\theta_l, q_l)$ we extract each (θ, q) in the list $L(p)$ of blocks which include p and hence which intersect B_{θ_l, q_l}. If $e(\theta, q) \neq 0$ then we set $e(\theta, q) = 0$. Since the selected blocks do not intersect, each $p \in \Gamma$ is covered by at most one selected blocks so over all iterations, this step requires $O(\sum_{p \in \Gamma} |L(p)|) = O(KS)$ operations. If $e(\theta, q) \neq 0$ then we also suppress $e(\theta, q)$ from the heap. Suppressing an element from a heap of size K requires $O(\log_2 K)$ operations. Since there are K elements in the heap after the initialization, the suppression of elements across all iterations is done with $O(K \log_2 K)$ operations.

Summing the operations of each steps, it results that $O(K(S^2 + \log_2 K))$ operations are sufficient over all iterations. If the operators $R_{\theta, q}$ are sparse operators then it reduces to $O(K(S + \log_2 K))$.
$\square$

In most applications, the geometry of blocks is highly structured, so as explained by the proof, the algorithm then only requires $O(K)$ memory. The computational upper bounds $O(K(S^2 + \log_2 K))$ and $O(K(S + \log_2 K))$ are pessimistic because they do not take into account the fact that signals have a sparse representation so blocks are not computed over regions where the frame coefficients have a negligible energy. The algorithm stops and does not select blocks covering all the frame indexes. The computational complexity is thus much smaller than with an $\mathbf{l}^1$ minimizer which requires $O(KS)$ operation per iteration.

VI. INTERPOLATIONS WITH SPARSE WAVELET MIXTURES

An adaptive directional image interpolation is computed by estimating sparse image mixture models in a wavelet frame. This section describes a fast block matching pursuit implementation, which requires $O(N \log N)$ operations and gives state-of-the-art interpolation results.

The subsampled image $y[n]$ for $n \in \mathcal{G}$ is decomposed in a translation invariant wavelet frame $\{\psi_{d,m}\}_{0 \leq d \leq 3, m \in \mathcal{G}}$ on a single scale (the finest one), and is reconstructed with a dual frame $\{\tilde{\psi}_{d,m}\}_{0 \leq d \leq 3, m \in \mathcal{G}}$ [40]. Wavelet coefficients are written

$$c(d, m) = \langle y, \psi_{d,m} \rangle \ .$$

The wavelet transform separates a low frequency image y_l projected over the low-frequency scaling filters $\{\psi_{0,m}\}_{m\in\mathcal{G}}$ and a high-frequency image y_h projected over the finest scale wavelets in three directions $\{\psi_{d,m}\}_{1\leq d\leq 3, m\in\mathcal{G}}$:

$$y_l = \sum_{m\in\mathcal{G}} c(0,m)\, \tilde{\psi}_{0,m} \text{ and } y_h = \sum_{d=1}^{3} c(d,m)\, \tilde{\psi}_{d,m}. \tag{28}$$

The low frequency image y_l has little aliasing and can thus be precisely interpolated with a cubic spline interpolator U^+. The high frequency image y_h is interpolated by selecting directional interpolators U_θ^+ for $\theta \in \Theta$, where Θ is a set of angles uniformly discretized between 0 and π. The appendix specifies the directional cubic spline interpolators U_θ^+ used in numerical experiments.

For each angle θ, a directional interpolator U_θ^+ is applied over a block $B_{\theta,q}$ of wavelet coefficients if the directional regularity factor $\|\bar{R}_{\theta,q}c\|$ is relatively small in the block. As explained in Section III, such a regularization is effective if the eigenvalues of $\bar{R}_{\theta,q}^* \bar{R}_{\theta,q}$ have an overall variation that is sufficiently large to distinguish regular from non-regular variations in the direction θ. This is obtained by choosing rectangular blocks $B_{\theta,q}$ that are elongated in the direction of θ. Each block $B_{\theta,q}$ in the spatial neighborhood of q is chosen to be identical in the three directions $d = 1,2,3$ so $1_{B_{\theta,q}}(d,m) = 1_{B_{\theta,q}}(m)$. Numerical experiments are performed with 20 angles θ, with blocks having a width of 2 pixels and a length between 6 and 12 pixels depending upon their orientation. Each block thus includes between 36 and 72 wavelet coefficients over the $d = 1,2,3$ directions.


d = 1　　　　**d = 2**　　　　**d = 3**
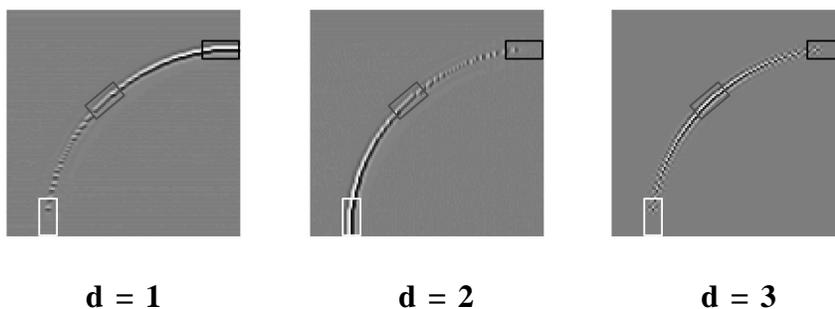
Fig. 1.　A block $B_{\theta,q}$ is composed of 3 elongated blocks (shown with the same gray level) of orientation θ in the 3 wavelet directions $d = 1,2,3$. In the neighborhood of an edge, a block is selected if the wavelet coefficients have regular variations in the direction θ, as shown by the 3 different blocks.

According to the algorithm of Section V, an adaptive interpolation estimator is obtained by estimating the mixing coefficients $\tilde{a}(\theta,q)$ of a mixture model which minimizes the Lagrangian
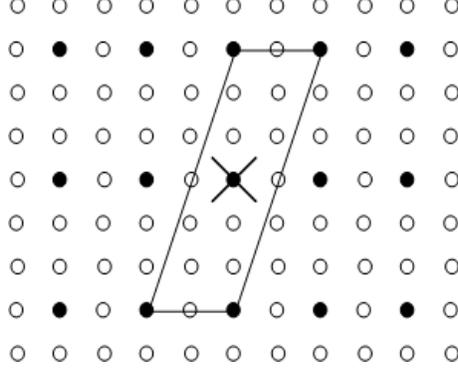
Fig. 2. Example of sampling grid Γ_θ (black dots) included in Γ (white dots) for $\theta = \arctan(3)$. A block $B_{\theta,q}$ of length $L = 6$ is shown, where $q \in \Gamma_\theta$ is indicated with a cross.

(20)

$$\mathcal{L}_1(\tilde{a}) \;=\; \frac{1}{2}\sum_{d=1}^{3}\sum_{m\in\mathcal{G}}|c(d,m)|^2\left(1-\sum_{\theta\in\Theta}\sum_{q\in\Gamma_\theta}\tilde{a}(\theta,q)\,1_{B_{\theta,q}}(m)\right)^2 \qquad (29)$$
$$+\lambda\sum_{\theta\in\Theta}\sum_{q\in\Gamma_\theta}|\tilde{a}(\theta,q)|\,\|\bar{R}_{\theta,q}c\|^2_{B_{\theta,q}}\,.$$

To further reduce computations, we do not implement exactly the regularity operators $\bar{R}_{\theta,q}$ corresponding to the interpolators U_θ^+ of the appendix. These operators are replaced by an approximation:

$$\bar{R}_{\theta,q}c(d,m) = c(d,m) - A_{\theta,q}c(d,m).$$

For each $d = 1,2,3$, $A_{\theta,q}c(d,m)$ is the average of the wavelet coefficients in the block $B_{\theta,q}$ which are located on the line of angle θ that goes through m. So $\bar{R}_{\theta,q}c$ can be computed with two operations per block point. The regularity norm $\|\bar{R}_{\theta,q}c\|^2_{B_{\theta,q}}$ is the energy of the coefficient variations relatively to their average in the direction θ. It is also the norm of the error when approximating wavelet coefficients by lines of constant wavelet coefficients along an angle θ in a block. These lines of wavelet coefficients are low-frequency bandlets of angle θ, as defined in [41]. The minimization in (29) can thus also be interpreted as an optimized approximation in orthogonal bandlets computed over adapted blocks of wavelet coefficients.

Block positions q are sampled along a grid Γ_θ to cover the image sampling grid: $\mathcal{G} = \cup_{q\in\Gamma_\theta}B_{\theta,q}$. If the block has a length L, the sampling grid is constructed by subsampling the image sampling grid by a factor $L/2$ in the horizontal or vertical direction closest to θ, and by

a factor 2 in the perpendicular direction, as illustrated in Figure 2. The resulting total number of blocks K over all angles is proportional to the number N of pixels, and in this implementation $K \leq 4N$.

Section V explains that the $\mathbf{l}^1$ Lagrangian (29) can be minimized with an iterative algorithm which requires $O(KS)$ operations per iteration, where $S = 72$ is the maximum block size. An orthogonal block matching pursuit requires much less operations. Since the directional regularity operators are sparse, Proposition 2 gives a heap implementation that requires $O(K(S + \log_2 K)) = O(N \log N)$ operations. Blocks are rectangles translated on a uniform grid, so block points and the intersection of blocks can be computed analytically with no storage requirements. The block matching pursuit implementation of Proposition 2 thus requires $O(K) = O(N)$ memory. To reduce grid effects, as explained at the end of Section V, several estimators are computed with different translations of these grids. The adaptive wavelet interpolator is derived from the averaged mixing coefficients (18).

Figure 3 shows an example of mixing coefficients $a_\theta(m)$ computed over wavelet coefficients along 20 angles θ of a discrete grid, which are the arctangent of rational numbers. The coefficient $a_\theta(m)$ are sparse and close to 1 only at the locations and in the appropriate direction θ where wavelet coefficients have a relatively large amplitude and are regular, which illustrates the accuracy of the direction estimation. Figure 4 also compares the energy of wavelets coefficients along all directions $\sum_{d=1}^{3} |c(d,m)|^2$ with the aggregation of mixing coefficients along all angles $\sum_{\theta \in \Theta} a_\theta(m)$. Observe that it is close to one where wavelet coefficients have a relatively large amplitude along geometric structures having some directional regularity. Directional interpolations are performed at these locations and are otherwise replaced by a separable interpolator.

The image low frequencies are interpolated with a cubic spline estimator U^+ and the highest frequency wavelets with the adaptive interpolator defined in (13):

$$\tilde{f} = U^+ y_l + \sum_{d=1}^{3} \sum_{m \in \Gamma} c(d,m) \left(\sum_{\theta \in \Theta} a_\theta(m) U_\theta^+ + a_r(m) U^+ \right) \tilde{\psi}_{d,m} , \tag{30}$$

with

$$a_\theta(m) = \sum_{q \in \Gamma_\theta} \tilde{a}(\theta,q) \mathbf{1}_{B_{\theta,q}}(m) \quad \text{and} \quad a_r(m) = 1 - \sum_{\theta \in \Theta} a_\theta(m) .$$

Since wavelets are translated, $\tilde{\psi}_{d,m}(n) = \tilde{\psi}_d(n-m)$, their interpolation are also translated:

$$U_\theta^+ \tilde{\psi}_{d,m}(n) = (U_\theta^+ \tilde{\psi}_d)(n-m) .$$
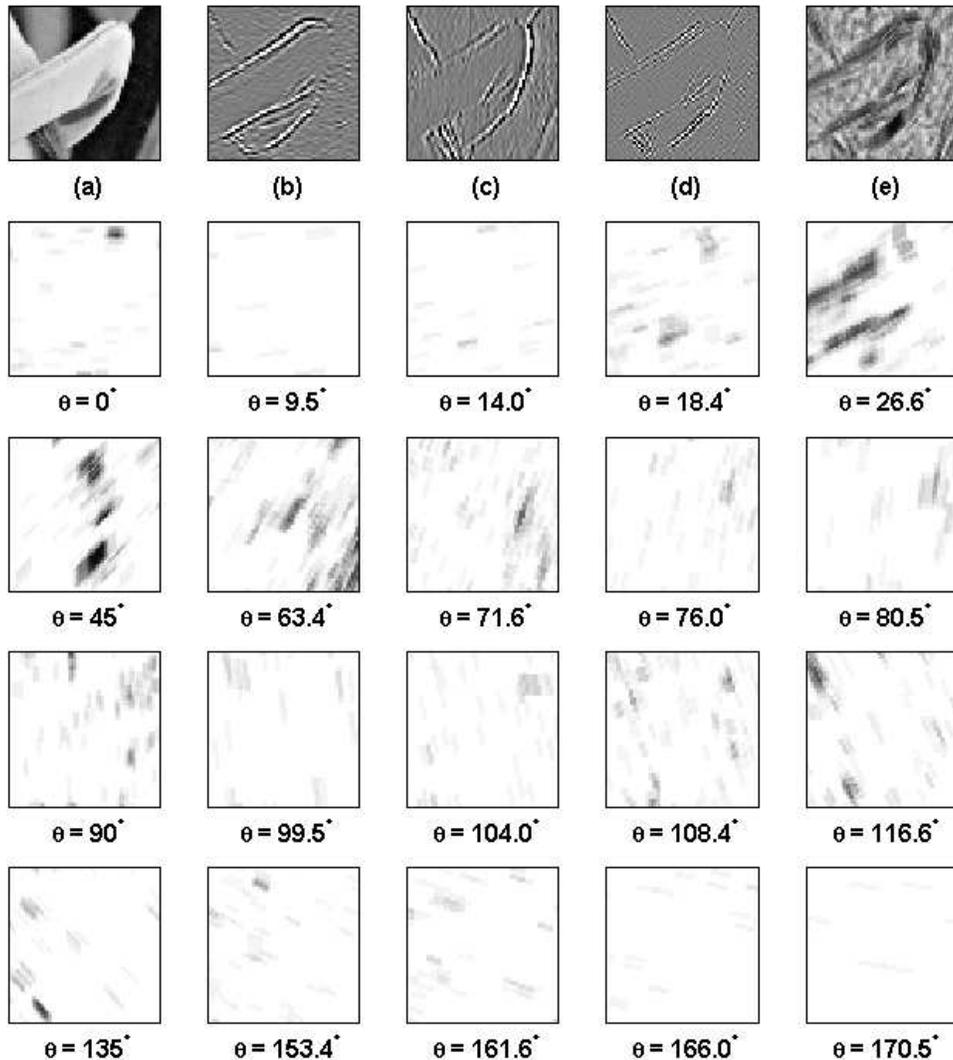
Fig. 3. (a) An image crop from Lena's hat. (b)-(d) Wavelet coefficients in the horizontal, vertical and diagonal directions. Black, gray and white represent negative, zero and positive coefficients. (e) Aggregation of mixing coefficients $\sum_{\theta \in \Theta} a_\theta(m)$. White mixing coefficients are close to 0 and black coefficients are close to 1. Second row to bottom row: each image gives the mixing coefficients $a_\theta(m)$ for a specific angle θ, which is the arctangent of a rational number.

If we precompute $\bar{\psi}_d = U^+ \tilde{\psi}_d$ and $\bar{\psi}_d^\theta = U_\theta^+ \tilde{\psi}_d$ then (30) is a reconstruction from an adapted set of wavelets

$$\tilde{f} = U^+ y_l + \sum_{d=1}^{3} \sum_{m \in \Gamma} c(d,m) \left(\sum_{\theta \in \Theta} a_\theta(m) \, \bar{\psi}_{d,m}^\theta + a_r(m) \, \bar{\psi}_{d,m} \right) . \quad (31)$$

For compactly supported wavelets, the wavelet interpolations are truncated to maintain a compact support. The mixing weights $a_\theta(m)$ are zero for most angles and (31) is computed with $O(N)$

Fig. 4. (a) Low-resolution Lena image y. (b) Energy of wavelet coefficients in all directions $\sum_{d=1}^{3} |c(d,m)|^2$. White and black pixels represent respectively small and large coefficients. (c) Aggregation of mixing coefficients $\sum_{\theta \in \Theta} a_\theta(m)$. White coefficients are close to 0 and black coefficients are close to 1.

operations. The overall adaptive interpolation algorithm is therefore implemented with $O(N \log N)$ operations.

The adaptive interpolator can also be computed by rewriting (30) as

$$\tilde{f} = U^+ \left(y_l + \sum_{d=1}^{3} \sum_{m \in \Gamma} a_r(m) \, c(d,m) \, \tilde{\psi}_{d,m} \right) + \sum_{\theta \in \Theta} U_\theta^+ \left(\sum_{d=1}^{3} \sum_{m \in \Gamma} a_\theta(m) \, c(d,m) \, \tilde{\psi}_{d,m} \right). \quad (32)$$

For each angle, an inverse wavelet transform is computed on wavelet coefficients multiplied by the mixing weights, and the resulting signal is interpolated in the corresponding direction.

The proposed image zooming algorithm, named hereafter SME (Sparse Mixing Estimation), is compared with a bicubic interpolation as well as recent super-resolution algorithms "NEDI" (New edge directed interpolation) [38], "DFDF" (Directional filtering and data fusion), "Curvelet" [26], "Contourlet" [45] and "SAI" (Soft-decision Adaptive Interpolation) [63]. As explained in Section I, NEDI, DFDF and SAI are adaptive directional interpolation methods. Curvelet and Contourlet are sparse inverse problem estimators described in Section II, computed in different dictionaries. Among previously published algorithms, SAI currently provides the best PSNR for spatial image interpolation [63]. All experiments are performed with softwares provided by the authors of these algorithms, and the SME software is available at *http://www.cmap.polytechnique.fr/~mallat/SME.htm*.

Figure 5 shows the six images used in the numerical experiments. Lena and Boat include both fine details and regular regions. Peppers and Cameraman are mainly composed of regular regions separated from sharp contours. Baboon is rich in fine details. Straws (from the Brodatz texture database) contains directional patterns that are superposed in various directions. These

high-resolution images are down-sampled by a factor 2×2. The resulting low-resolution images are then zoomed by the algorithms under comparison.
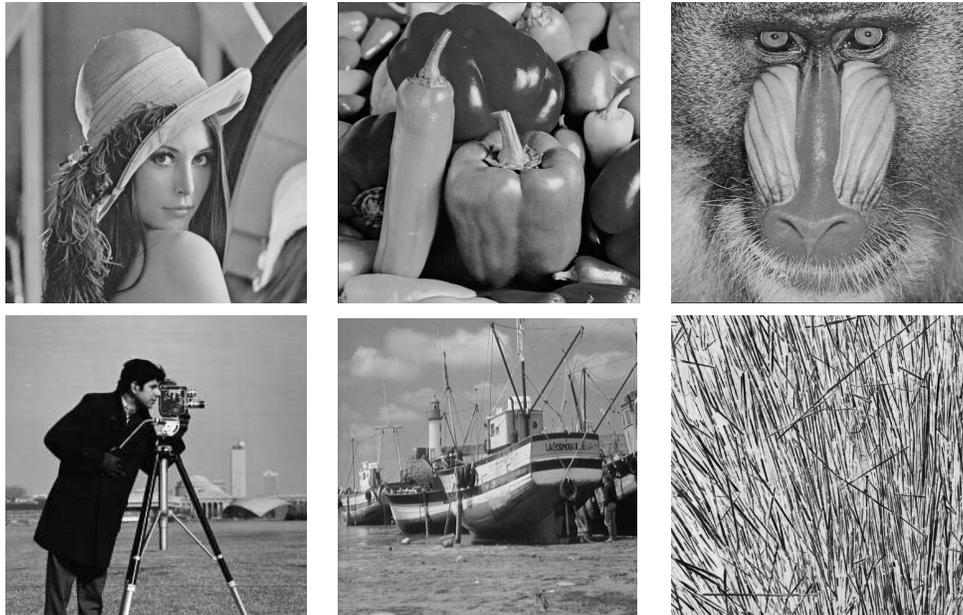
Fig. 5.   Images used in the numerical experiments. From top to bottom, left to right: Lena, Peppers, Baboon, Cameraman, Boat, Straws.

Table I gives the PSNRs generated by all algorithms for the images in Figure 5. The SME algorithm is implemented with a Lagrangian multiplier $\lambda = 0.6$ in (29). For all these images, the results obtained with an orthogonal matching pursuit minimization of the Lagrangian or with an iterative l^1 minimizer are within 0.1db. In the following, all SME numerical results are thus computed with an orthogonal block matching pursuit which requires much less operations. SME and SAI give similar PSNRs for all the images, the overall gain of SME being slightly better. Their gain in PSNR is significantly larger than with all other algorithms.

The sparse Contourlet interpolation algorithm yields almost the same PSNR as a bicubic interpolation but often provides better image quality. Sparse estimations in a curvelet dictionary as implemented in [26] provides good results for image inpainting but is not suitable for image zooming.

Figure 6 compares the interpolated image obtained by different algorithms. The local PSNRs on the cropped images are reported as well. Bicubic interpolations produce some blur and jaggy artifacts in the zoomed images. These artifacts are reduced to some extent by the NEDI and

	Bicubic	NEDI	DFDF	Curvelet	Contourlet	SAI	SME
Lena	33.93	33.77	33.91	24.31	33.92	**34.68**	34.58
Peppers	32.83	33.00	33.18	23.60	33.10	**33.52**	**33.52**
Baboon	22.92	23.16	22.83	20.34	22.53	**23.19**	23.16
Cameraman	25.37	25.42	25.67	19.50	25.35	25.88	**26.26**
Boat	29.24	29.19	29.32	22.20	29.25	29.68	**29.76**
Straws	20.53	20.54	20.70	17.09	20.52	21.48	**21.61**
Ave. gain	0	0.04	0.13	-6.30	-0.02	0.60	**0.68**

TABLE I

COMPARISON OF IMAGE ZOOMING ALGORITHMS. PSNRS (IN DB) ARE COMPUTED OVER IMAGES OF FIGURE 5. FROM LEFT TO RIGHT: BICUBIC INTERPOLATION, NEDI [38], DFDF [62], CURVELET [26], CONTOURLET [45], SAI [63] AND SME (SPARSE MIXING ESTIMATOR). THE BOTTOM ROW SHOWS THE AVERAGE GAIN OF EACH METHOD RELATIVE TO BICUBIC INTERPOLATION. THE HIGHEST PSNR IN EACH ROW IS IN BOLDFACE.

DFDF algorithms, but the image quality is lower than with SME and SAI algorithms, as shown by the PSNRs. The SME algorithms restores slightly better regular geometrical structures than SAI, as shown by the middle leg of the tripod in Cameraman and the beards of Baboon. The contourlet algorithm is able to restore the geometrical structures (see Baboon's beard) when the underlying contourlet vectors are accurately estimated. However, as explained in Section II, the vector support recovery is not stable. When the approximating contourlet vectors are not estimated correctly, it produces directional artifact patterns, which offsets its gain in PSNR.

Figure 7 further compares SME with bicubic interpolation. SME improves the PSNR and the visual image quality where the image has some directional regularity. It appears in the straws, the hat border and the hairs of various directions. Otherwise, it is similar to a bicubic interpolation since it also implements a non-directional separable interpolation.

VII. CONCLUSION

This paper introduces a new class of adaptive estimators obtained by mixing a family of linear inverse estimators, derived from different priors on the signal regularity. Mixing coefficients are calculated in a frame over blocks of coefficients having an appropriate regularity and providing a sparse signal representation. They are computed by minimizing an $\mathbf{l}^1$ norm which is weighted by the signal regularity in each block. This regularization improves the estimation for highly unstable inverse problems relatively to lasso estimators which compute an $\mathbf{l}^1$ norm or a mixed $\mathbf{l}^2$ and $\mathbf{l}^1$ norm over blocks of dictionary coefficients. A fast orthogonal matching pursuit algorithm

High-resolution image   **Low-resolution image**   **Bicubic 21.60 dB**   **SME 23.54 dB**

NEDI 22.00 dB   **DFDF 22.18 dB**   **Contourlet 21.79 dB**   **SAI 22.38 dB**

High-resolution image   **Low-resolution image**   **Bicubic 23.00 dB**   **SME 23.88 dB**

NEDI 23.80 dB   **DFDF 23.16 dB**   **Contourlet 23.06 dB**   **SAI 23.62 dB**
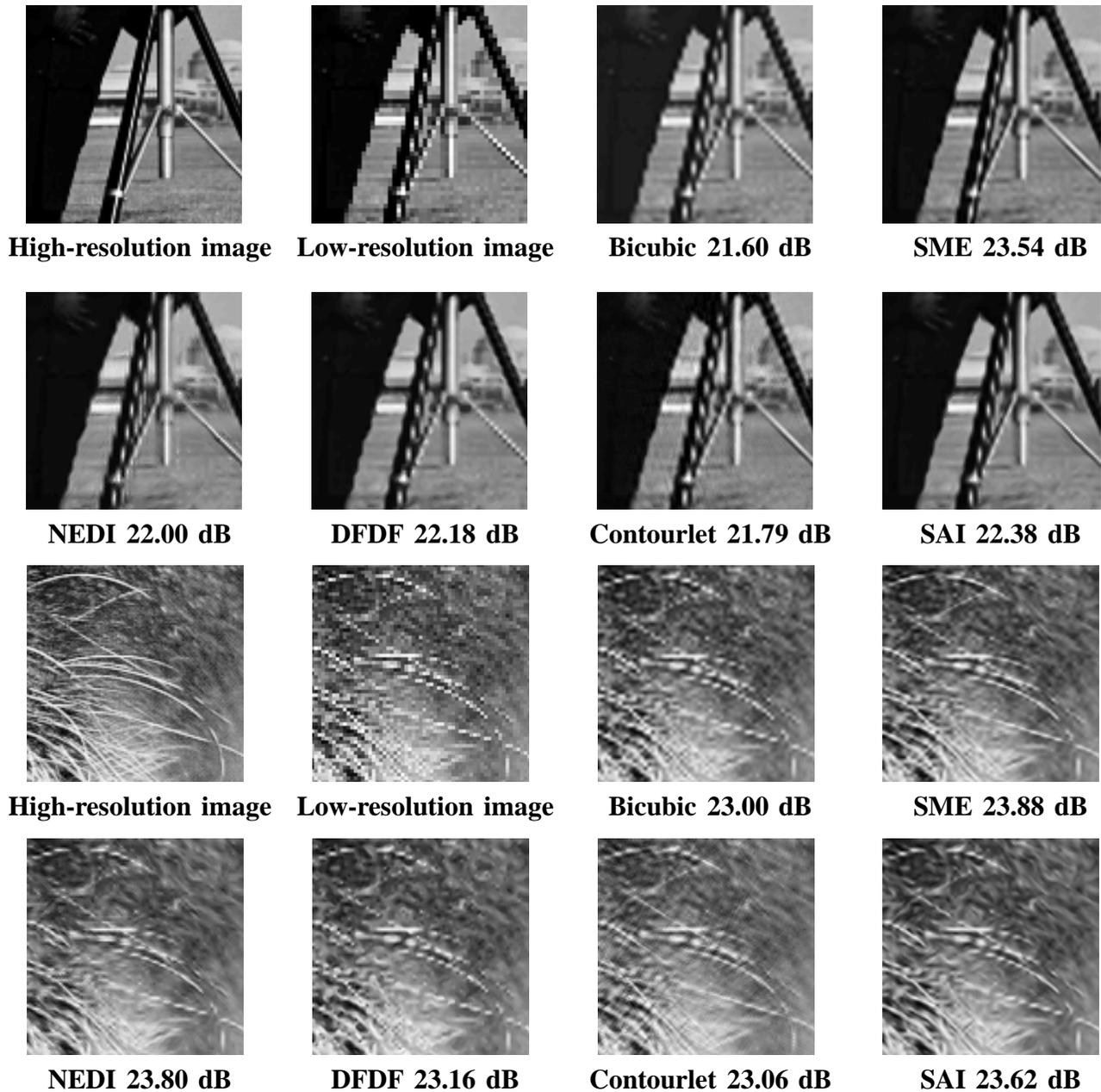
Fig. 6. PSNRs (in dB) are computed in the cropped images (from Cameraman and Babboon). From left to right: high-resolution image, low-resolution image (shown at the same scale by enlarging the pixel size), bicubic interpolation, SME (Sparse Mixing Estimator), NEDI [38], DFDF [62], Contourlet [45], SAI [63].
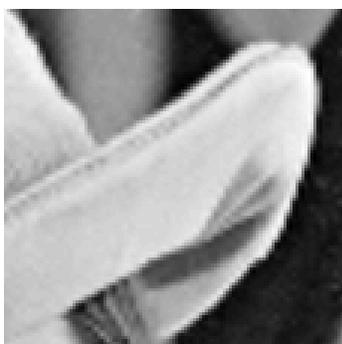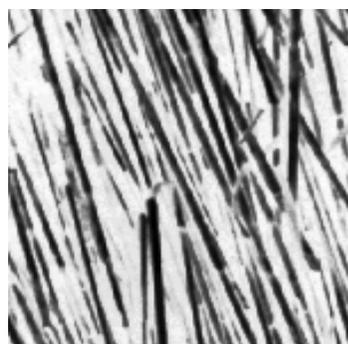
Fig. 7. PSNRs (in dB) are computed in the illustrated cropped images (from Lena and Straws). From left to right: high-resolution image, bicubic interpolation, SME.

is introduced to reduce the number of operations. A particular application to image interpolations is studied by mixing directional interpolators over oriented blocks in a wavelet frame. For an image of N pixels, the computational complexity is $O(N \log N)$ and it provides state-of-the-art interpolation results.

APPENDIX

A cubic-spline [57] directional interpolator U_θ^+ in a direction θ is described for upscaling images by a factor 2. It is illustrated in Fig. 8. Original samples are shown as crosses.
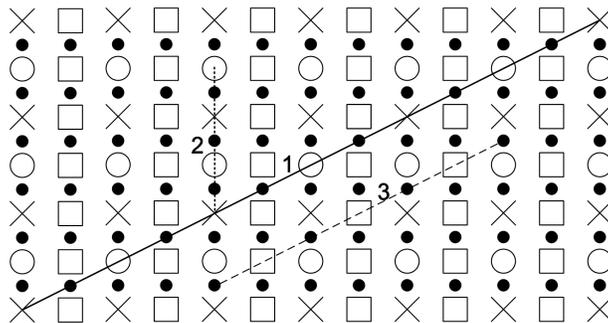

Fig. 8. Directional interpolation scheme. Samples represented by the crosses, circles, dots and squares will be named crosses, circles, dots and squares for short. The directional interpolation starts from the low-resolution image defined on the crosses and proceeds in three steps: (1) One-dimensional interpolations along the angle θ, which reconstructs the circles from the crosses. (2) A one-dimensional vertical interpolation which reconstructs the dots from the crosses and the circles. The dots do not belong to the resulting high-resolution image and will be used in the following step. (3) Another one-dimensional interpolation along θ, which reconstructs the squares from the dots.

- Step 1 computes a one-dimensional interpolations in the direction θ. We consider all lines of angle θ that intersect original image samples (crosses in Fig. 8) and we compute mid-points (circles) between image samples (crosses), with a cubic spline interpolation. This operation oversamples by a factor two either the image rows, or the image columns, or the diagonals of angle $\pm\pi/4$. The missing coefficients are shown as squares in Fig. 8.
- Step 2 computes new samples (dots) with a cubic spline interpolation along these oversampled rows, columns or diagonals. This interpolation introduces little aliasing because of the oversampling provided by the previous step. The positions of these new samples (dots) are chosen so that any missing coefficient (square) is a mid-point between two dots on a line of angle θ.
- Step 3 computes missing samples (squares) with a cubic spline linear interpolation along the direction θ from the previously calculated new samples (dots).

REFERENCES

[1] V.R. Algazi, G.E. Ford, and R. Potharlanka. Directional interpolation of images based on visual properties and rank order filtering. In *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, volume 4, pages 3005–3008, 1991.

[2] J. Allebach and P.W. Wong. Edge-directed interpolation. In *Image Processing, 1996. Proceedings., International Conference on*, volume 3, 1996.

[3] H. A. Aly and E. Dubois. Image up-sampling using total-variation regularization with a new observation model. *IEEE Transactions on Image Processing*, 14(10):1647–1659, 2005.

[4] F. Bach. Consistency of the group Lasso and multiple kernel learning. *The Journal of Machine Learning Research*, 9:1179–1225, 2008.

[5] R.G. Baraniuk, V. Cevher, M.F. Duarte, and C. Hegde. Model-based compressive sensing. *preprint*, 2008.

[6] S. D. Bayrakeri and R. M. Mersereau. A new method for directional image interpolation. In *Acoustics, Speech, and Signal Processing, 1995. ICASSP-95., 1995 International Conference on*, volume 4, 1995.

[7] A. Biancardi, L. Cinque, and L. Lombardi. Improvements to image magnification. *Pattern Recognition*, 35(3):677–687, 2002.

[8] P. Bofill and M. Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353–2362, 2001.

[9] D. Calle and A. Montanvert. Super-resolution inducing of an image. In *Proc. IEEE Int. Conf. Image Processing*, volume 3, pages 232–235, 1998.

[10] E.J. Candes and D.L. Donoho. New tight frames of curvelets and optimal representations of objects with C 2 singularities. *Comm. Pure Appl. Math*, 56:219–266, 2004.

[11] E.J. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transactions on Information Theory*, 52(12):5406–5425, 2006.

[12] W.K. Carey, D.B. Chuang, and S.S. Hemami. Regularity-preserving image interpolation. *IEEE transactions on image processing*, 8(9):1293–1297, 1999.

[13] S. Carrato, G. Ramponi, and S. Marsi. A simple edge-sensitive image interpolation filter. In *IEEE International Conference on Image Processing*, volume 3, 1996.

[14] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1999.

[15] T. Chen, H. R. Wu, and B. Qiu. Image interpolation using across-scale pixel correlation. In *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01)*, volume 3, 2001.

[16] M.A. Chughtai and N. Khattak. An Edge Preserving Locally Adaptive Anti-aliasing Zooming Algorithm with Diffused Interpolation. In *Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision*. IEEE Computer Society Washington, DC, USA, 2006.

[17] T.H. Corman, C.E. Leiserson, R.L. Rivest, and C. Stein. Introduction to algorithms, 1990.

[18] A.M. Darwish, M.S. Bedair, and S.I. Shaheen. Adaptive resampling algorithm for image zooming. *IEE Proceedings-Vision, Image and Signal Processing*, 144(4):207–212, 1997.

[19] I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Comm. Pure Appl. Math*, 57:1413–1457, 2004.

[20] M.N. Do and M. Vetterli. The contourlet transform: an efficient directional multiresolution image representation. *IEEE Transactions on image processing*, 14(12):2091–2106, 2005.

[21] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[22] M. Elad and D. Datsenko. Example-based regularization deployed to super-resolution reconstruction of a single image. *The Computer Journal*, 2007.

[23] M. Elad, J.L. Starck, P. Querre, and D.L. Donoho. Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). *Appl. Comput. Harmon. Anal*, 19:340–358, 2005.

[24] Y.C. Eldar, P. Kuppinger, and H. Bölcskei. Compressed Sensing of Block-Sparse Signals: Uncertainty Relations and Efficient Recovery. *arXiv: 0906.3173, submitted to IEEE Transactions on Signal Processing*, 2009.

[25] Y.C. Eldar and M. Mishali. Robust recovery of signals from a union of subspaces. *arXiv.org 0807.4581, to appear in IEEE Trans. Inform. Theory*, 2008.

[26] M.J. Fadili, J.L. Starck, and F. Murtagh. Inpainting and Zooming Using Sparse Representations. *The Computer Journal*, 2007.

[27] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.

[28] W.T. Freeman, T.R. Jones, and E.C. Pasztor. Example-Based Super-Resolution. *IEEE Computer Graphics and Applications*, pages 56–65, 2002.

[29] J. Huang, T. Zhang, and D. Metaxas. Learning with Structured Sparsity. *arXiv:0903.3002v1*, 2009.

[30] M. Irani and S. Peleg. Super resolution from image sequences. In *Pattern Recognition, Proceedings., 10th International Conference on*, volume 2, 1990.

[31] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal of Visual Communication and Image Representation*, 4(4), 1993.

[32] R. Jenatton, J.Y. Audibert, and F. Bach. Structured Variable Selection with Sparsity-Inducing Norms. *arXiv:0904.3523v1*, 2009.

[33] K. Jensen and D. Anastassiou. Subpixel edge localization and the interpolation of still images. *IEEE Transactions on Image Processing*, 4(3):285–295, 1995.

[34] H. Jiang and C. Moloney. A new direction adaptive scheme for image interpolation. In *International Conference on Image Processing*, volume 3, pages 369–372, 2002.

[35] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 29(6):1153–1160, 1981.

[36] S.W. Lee and J. K. Paik. Image interpolation using adaptive fast B-spline filtering. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, 1993.

[37] G. Leung and A.R. Barron. Information theory and mixing least-squares regressions. *IEEE Transactions on Information Theory*, 52(8):3396–3410, 2006.

[38] X. Li and M. T. Orchard. New edge-directed interpolation. *Image Processing, IEEE Transactions on*, 10(10):1521–1527, 2001.

[39] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM Journal on Numerical Analysis*, pages 1–37, 2002.

[40] S. Mallat. *A Wavelet Tour of Signal Processing: The Sparse Way, 3rd edition*. Academic Press, 2008.

[41] S. Mallat and G. Peyré. Orthogonal bandlet bases for geometric images approximation. *Comm. Pure Appl. Math*, 61(9):1173–1212, 2008.

[42] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[43] S. Masnou and J.M. Morel. Level lines based disocclusion. In *International Conference on Image Processing*, pages 259–263, 1998.

[44] B. S. Morse and D. Schwartzwald. Image magnification using level-set reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, 2001.

[45] N. Mueller, Y. Lu, and M.N. Do. Image interpolation using multiscale geometric representations. In *Computational Imaging V. Edited by Bouman, Charles A.; Miller, Eric L.; Pollak, Ilya. Proceedings of the SPIE*, volume 6498, page 64980A, 2007.

[46] S.C. Park, M.K. Park, and M.G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE Signal Processing Magazine*, 20(3):21–36, 2003.

[47] J. Portilla and E. Simoncelli. Image restoration using gaussian scale mixtures in the wavelet domain. In *In Proc. IEEE ICIP*, pages 965–968, 2003.

[48] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Transactions on Image Processing*, 2008.

[49] K. Ratakonda and N. Ahuja. POCS based adaptive image magnification. In *International Conference on Image Processing*, pages 203–207, 1998.

[50] V. Roth and B. Fischer. The Group-Lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the 25th international conference on Machine learning*, pages 848–855. ACM New York, NY, USA, 2008.

[51] H. Shi and R. Ward. Canny edge based image expansion. In *IEEE International Symposium on Circuits and Systems*, volume 1, 2002.

[52] M. Stojnic, F. Parvaresh, and B. Hassibi. On the Reconstruction of Block-Sparse Signals With an Optimal Number of Measurements. *IEEE Transactions on Signal Processing*, 57(2):3075–3085, 2009.

[53] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.

[54] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. In *Soviet Math. Dokl*, volume 4, pages 1035–1038, 1963.

[55] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, 2004.

[56] J. A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.

[57] M. Unser. Splines: A perfect fit for signal and image processing. *IEEE Signal processing magazine*, 16(6):22–38, 1999.

[58] J. D. Van Ouwerkerk. Image super-resolution survey. *Image and Vision Computing*, 24(10):1039–1052, 2006.

[59] Q. Wang and R. K. Ward. A New Orientation-Adaptive Interpolation Method. *Image Processing, IEEE Transactions on*, 16(4):889–900, 2007.

[60] J. Yang, J. Wright, T. Huang, and Y. Ma. Image super-resolution as sparse representation of raw image patches. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pages 1–8, 2008.

[61] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, 68(1):49–67, 2006.

[62] L. Zhang and X. Wu. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Transactions on Image Processing*, 15(8):2226, 2006.

[63] X. Zhang and X. Wu. Image interpolation by adaptive 2-d autoregressive modeling and soft-decision estimation. *IEEE Transactions on Image Processing*, 17(6):887–896, 2008.

[64] M. Zibulevsky and B. A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.

[65] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B*, 67(2):301–320, 2005.