Inpainting and zooming using sparse representations

Jalal M. Fadili, Jean-Luc Starck, Fionn Murtagh

The Computer Journal Advance Access published July 29, 2007

doi:10.1093/comjnl/bxm055

Inpainting and Zooming Using Sparse Representations

M.J. FADILI[1], J.-L. STARCK[2] AND F. MURTAGH[3,*]

[1]*GREYC CNRS UMR 6072, Image Processing Group, ENSICAEN 14050, Caen Cedex, France*
[2]*CEA-Saclay, DAPNIA/SEDI-SAP, Service d'Astrophysique, F-91191 Gif sur Yvette, France*
[3]*Department of Computer Science, Royal Holloway, University of London, Egham TW20 0EX, UK*
**Corresponding author: fionn@cs.rhul.ac.uk*

Representing the image to be inpainted in an appropriate sparse representation dictionary, and combining elements from Bayesian statistics and modern harmonic analysis, we introduce an expectation maximization (EM) algorithm for image inpainting and interpolation. From a statistical point of view, the inpainting/interpolation can be viewed as an estimation problem with missing data. Toward this goal, we propose the idea of using the EM mechanism in a Bayesian framework, where a sparsity promoting prior penalty is imposed on the reconstructed coefficients. The EM framework gives a principled way to establish formally the idea that missing samples can be recovered/interpolated based on sparse representations. We first introduce an easy and efficient sparse-representation-based iterative algorithm for image inpainting. Additionally, we derive its theoretical convergence properties. Compared to its competitors, this algorithm allows a high degree of flexibility to recover different structural components in the image (piecewise smooth, curvilinear, texture, etc.). We also suggest some guidelines to automatically tune the regularization parameter.

Keywords: EM algorithm; sparse representations; inpainting; interpolation; penalized likelihood

Received 30 June 2006; revised 15 February 2007

1. INTRODUCTION

Inpainting is to restore missing image information based upon the still available (observed) cues from destroyed, occluded or deliberately masked subregions of the images. The keys to successful inpainting are to infer robustly the lost information from the observed cues. The inpainting can also be viewed as an interpolation or a disocclusion problem. The classical image inpainting problem can be stated as follows. Suppose the ideal complete image is $\mathbf{x}$ defined on a finite domain Ω (the plane), and its degraded version (but not completely observed) $\mathbf{y}$. The observed (incomplete) image $\mathbf{y}_{\text{obs}}$ is the result of applying the lossy operator $\mathcal{M}$ on $\mathbf{y}$

$$\mathcal{M} : \mathbf{y} \mapsto \mathbf{y}_{\text{obs}} = \mathcal{M}[\mathbf{y}] = \mathcal{M}[\mathbf{x} \odot \varepsilon] \tag{1}$$

where $\odot$ is any composition of two arguments (e.g. '+' for additive noise, etc.), ε is the noise. $\mathcal{M}$ is defined on $\Omega \backslash E$, where E is a Borel measurable set. A typical example of $\mathcal{M}$ that will be used throughout this paper is the binary mask; a diagonal matrix with ones (observed pixel) or zeros (missing pixel), and $\mathbf{y}_{\text{obs}}$ is a masked image with zeros wherever a pixel is missing. Inpainting is to recover $\mathbf{x}$ from $\mathbf{y}_{\text{obs}}$, which is an inverse ill-posed problem.

1.1. State of affairs

Non-texture image inpainting has received considerable interest and excitement since the pioneering paper by Masnou and Morel [1, 2] who proposed variational principles for image disocclusion (the term inpainting was not used yet). A recent wave of interest in inpainting has also started from the paper of [3], where applications in the movie industry, video and art restoration were unified. These authors proposed nonlinear partial differential equations (PDE) for non-texture inpainting. Basically, the major mechanisms to get information into the inpainting domain are:

(i) Diffusion and transport PDE/Variational principle: in [3–6], the PDE is obtained phenomenologically, and axiomatically in [7]. Ballester *et al.* in a series of papers [8–10], proposed to solve the inpainting problem by considering a joint interpolation of the orthogonal direction of level lines and of the gray levels. In [11], the authors proposed a curvature-driven

diffusion image inpainting method. Chan and Shen [12] systematically investigated inpainting based on the Bayesian and (possibly hybrid) variational principles with different penalizations (TV, ℓ_1 norm on wavelet coefficients). By functionalizing Euler's elastica energy, Masnou and Morel [1, 2], proposed an elastica-based variational inpainting model, and [13] developed its mathematical foundation and a PDE computational scheme. Following in the footsteps of Chan and Shen, authors in [1] adapted the Mumford–Shah image segmentation model to the inpainting problem. Chan *et al.* [15] also employed a TV variational approach for joint image inpainting and deblurring. They have recently applied this work to superresolution [15].

There are other recent references concerned with PDE/variational inpainting and the interested reader may refer for example to the UCLA CAM reports website[1] for an abundant literature (e.g. [17, 18], etc.). For an extended survey of the mathematics involved in the above frameworks and some applications, see [19, Chapter 6, 20]. Note however that the main intrinsic limitations of these approaches is that they cannot deal with texture inpainting. They are suitable for geometrical parts of the image and perform well only for non-texture inpainting. Thus, texture and large empty gaps can be a major issue.

(ii) Exemplar region fill-in: the basic idea behind these algorithms is to fill-in the missing regions with available information from some best matching candidates (e.g. surrounding pixels/regions [21, 22], best correspondence maps [23]). The goal is to synthesize a complete, visually plausible and coherent image, where the visible parts of the image serve as a reference/training set to infer the unknown parts. For example, in [21], the authors proposed a Markov random field-based synthesis approach only suitable for texture inpainting. Bertalmìo and co-workers [22, 24] adapted Meyer's $u + v$ model to decompose the image into its natural (cartoon) and texture parts, then inpaint the natural part using a PDE-based method and a fill-in-based approach for the texture part following the work of [21]. Finally, the inpainted parts are recombined. Other authors have also proposed to first restore the geometrical (or natural) part combined with a fill-in texture synthesis method [25, 26]. In work by Criminisi *et al.* [27], the authors proposed an algorithm that allows one to fill-in an image progressively by looking for best-matching prototypes in a dictionary built from a set of patterns around the missing subregions. Drori *et al.* [28] also proposed a fill-in image completion method which bears similarities with [27], although the fill-in step was differently approached.

(iii) Compressed/ive Sensing: more recently, [29] introduced a novel inpainting algorithm that is capable of reconstructing both texture and cartoon image contents. This algorithm is a direct extension of the Morphological Component Analysis (MCA), designed for the separation of an image into different semantic components [30, 31]. The arguments supporting this method were borrowed from the very exciting and recent theory of Compressed/ive Sensing recently developed by Donoho [32], Donoho and Tsaig [33] and Candès *et al.* [34–36].

[1] http://www.math.ucla.edu/applied/cam/index.html

1.2. This paper

Combining elements from statistics and modern harmonic analysis theories, we here introduce an EM algorithm [37, 38] for image inpainting/interpolation based on a penalized maximum-likelihood formulated using linear sparse representations, i.e. $\mathbf{x} = \boldsymbol{\Phi}\boldsymbol{\alpha}$, where the image $\mathbf{x}$ is supposed to be efficiently represented by a few atoms in the dictionary. Therefore, a sparsity promoting prior penalty is imposed on the reconstructed coefficients. From a statistical point of view, the inpainting can be viewed as an estimation problem with incomplete or missing data, where the EM framework is a very general tool in such situations. The EM algorithm formalizes the idea of replacing the missing data by estimated ones from coefficients of the previous iteration, and then we reestimate the new expansion coefficients from the complete formed data, and iterate the process until convergence. We here restrict ourselves to zero-mean additive white Gaussian noise (AWGN), even if the theory of the EM can be developed for the regular exponential family. The EM framework gives a principled way to establish formally the idea that missing samples can be recovered based on sparse representations. Furthermore, owing to its well-known theoretical properties, the EM algorithm allows us to investigate the convergence behavior of the inpainting algorithm. Some results are shown to illustrate our algorithm on inpainting, zooming and disocclusion tasks.

This paper is organized as follows. In Section 2, we will introduce the necessary ingredients of sparse representations and concepts from modern harmonic analysis. In Section 3, we state the inpainting problem, give the necessary material about the EM mechanism and Bayesian estimation, and the inpainting algorithm is also introduced. Section 4 states some key convergence analysis results whose proofs are outlined in the Appendix, and discusses the influence of the regularization parameter. Section 5 is devoted to presenting and discussing experimental results including inpainting and zooming. We conclude with a brief summary of our work and some perspectives on its likely future development.

1.3. Notation

Let $\mathcal{H}$ a real Hilbert space, which is in our finite-dimensional case, a subset of the vector space $\mathbb{R}^n$. We denote by $\|.\|_2$ the norm associated with the scalar product in $\mathcal{H}$, and $\mathbf{I}$ is the identity operator on $\mathcal{H}$. The $\mathbf{x}$ and $\boldsymbol{\alpha}$ are, respectively, reordered vectors of image samples and transform coefficients. Let f: $\mathcal{H} \mapsto \mathbb{R} \cup \{\infty\}$, where the domain of f is dom $f = \{\mathbf{x} \in \mathcal{H} \mid f(\mathbf{x}) < \infty\}$. A function is proper if its domain is non-empty, i.e. not identically $+\infty$ everywhere. A function f is coercive, if $\lim_{\|\mathbf{x}\|_2 \to +\infty} f(\mathbf{x}) = +\infty$.

2. SPARSE IMAGE REPRESENTATION

2.1. Sparse representations

Suppose $\mathbf{x} \in \mathcal{H}$. An $\sqrt{n} \times \sqrt{n}$ image $\mathbf{x}$ can be written as the superposition of (a few) elementary functions $\phi_\gamma(s)$ (atoms) parameterized by $\gamma \in \Gamma$ such that (Γ is countable)

$$\mathbf{x}(s) = \sum_{\gamma \in \Gamma} \boldsymbol{\alpha}_\gamma \phi_\gamma(s), \quad \text{Card}\,\Gamma = L. \tag{2}$$

Popular examples of Γ include: frequency (Fourier), scale-translation (wavelets), scale-translation-frequency (wavelet packets), translation-duration-frequency (cosine packets), scale-translation-angle (e.g. curvelets, bandlets, contourlets, wedgelets, etc.). The dictionary $\boldsymbol{\Phi}$ is the $n \times L$ matrix whose columns are the generating atoms $\{\phi_\gamma\}_{\gamma \in \Gamma}$, which are supposed to be normalized to a unit ℓ_2-norm. The forward transform is defined via a non-necessarily square full rank matrix $\mathbf{T} = \boldsymbol{\Phi}^{\mathrm{T}} \in \mathbb{R}^{L \times n}$, with $L \geq N$. When $L > n$ the dictionary is said to be redundant or overcomplete.

Owing to recent advances in modern harmonic analysis, many redundant systems, like the undecimated wavelet transform, curvelet, contourlet, steerable or complex wavelet pyramids, were shown to be very effective in sparsely representing images. By sparsity, we mean that we are seeking a good representation of $\mathbf{x}$ with only very few non-zero coefficients, i.e. $\|\boldsymbol{\alpha}\|_0 \ll n$. In most practical situations, the dictionary is built by taking the union of one or several (sufficiently incoherent) transforms, generally each corresponding to an orthogonal basis or a tight frame. In the case of a simple orthogonal basis, the inverse transform is trivially $\boldsymbol{\Phi} = \mathbf{T}^{\mathrm{T}}$; whereas assuming that $\mathbf{T}$ is a tight frame implies that the frame operator satisfies $\mathbf{T}^{\mathrm{T}}\mathbf{T} = c\mathbf{I}$, where $c > 0$ is the tight frame constant. Hence, $\mathbf{T}^{\mathrm{T}} = \boldsymbol{\Phi}$ is the Moore–Penrose pseudo-inverse transform (corresponding to the minimal dual synthesis frame), up to the constant c. In other words, the pseudo-inverse reconstruction operator $\mathbf{T}^{+}$ reduces to $c^{-1}\boldsymbol{\Phi}$.

2.2. Choosing a dictionary

Choosing an appropriate dictionary is a key step toward a good sparse representation, hence inpainting and interpolation. A core idea here is the concept of morphological diversity, as initiated in [30]. When the transforms are amalgamated in one dictionary, they have to be chosen such that each leads to sparse representations over the parts of the images it is serving. Thus, for example (see Fig. 1), to represent efficiently isotropic structures in an image, a qualifying choice is the wavelet transform [39, 40]. The curvelet system [41, 42] is a very good candidate for representing piecewise smooth (C^2) images away from C^2 contours. The ridgelet transform [43, 44] has been shown to be very effective for representing global lines in an image. The local discrete cosine transform (DCT) [39] is well suited to represent locally stationary textures. These transforms are also computationally tractable particularly in large-scale applications, and never explicitly implementing $\boldsymbol{\Phi}$ and $\mathbf{T}$. The associated implicit fast analysis and synthesis operators have typically complexities of $O(n)$ (e.g. orthogonal or bi-orthogonal wavelet transform) or $O(n \log n)$ (e.g. ridgelet, curvelet, local DCT transforms). Another desirable requirement that the merged transforms have to satisfy is that when a transform sparsely represents a part in the image, it yields non-sparse representations on the other content type. This is intimately related to the concept of mutual incoherence, which plays a fundamental role in sparse solutions of overcomplete (redundant) linear systems. For a given dictionary, it is defined as the maximum in magnitude of the off-diagonal elements of its Gram matrix (see e.g. [45–48] for more details).

3. PENALIZED MAXIMUM-LIKELIHOOD ESTIMATOR WITH MISSING DATA

3.1. Penalized maximum-likelihood estimator estimation

First, let's ignore the missing data mechanism and consider the complete n-dimensional observed vector (by simple reordering) $\mathbf{y}$. Adopting a GLM framework, we assume that $\mathbf{y}$ has a density in the classical exponential family, taking the form [49]

$$f_{\mathbf{Y}}(\mathbf{y};\, \boldsymbol{\eta},\, \sigma) = h(\mathbf{y},\, \sigma) \exp\left[\frac{\boldsymbol{\eta}^{\mathrm{T}}\mathbf{y} - A(\boldsymbol{\eta})}{g(\sigma)}\right] \tag{3}$$

for some specific functions $A(.)$, $g(.)$ and $h(.)$, $\boldsymbol{\eta}$ is the canonical parameter and σ is the scale parameter. From classical results on exponential families [49], the mean of $\mathbf{Y}$ is related to $\boldsymbol{\eta}$ via $\mathsf{E}[\mathbf{Y}_i] = a'(\boldsymbol{\eta}_i)$, because typically, $A(\boldsymbol{\eta}) = \sum_{i=1}^{n} a(\boldsymbol{\eta}_i)$. There also exists a one-to-one transform, called the link function, relating $\mathsf{E}[\mathbf{Y}]$ to $\boldsymbol{\Phi}\boldsymbol{\alpha}$. The most important case corresponds to the *canonical* link function $(a')^{-1}(.)$, or $\boldsymbol{\eta} = \boldsymbol{\Phi}\boldsymbol{\alpha}$, that is $(\boldsymbol{\Phi}\boldsymbol{\alpha})_i = (a')^{-1}(\mathsf{E}[\mathbf{Y}_i])$. In this case, $\boldsymbol{\alpha}$ can be viewed as parameters of a canonical sub-family of the original exponential family with sufficient statistic $\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y}$. For instance, the canonical

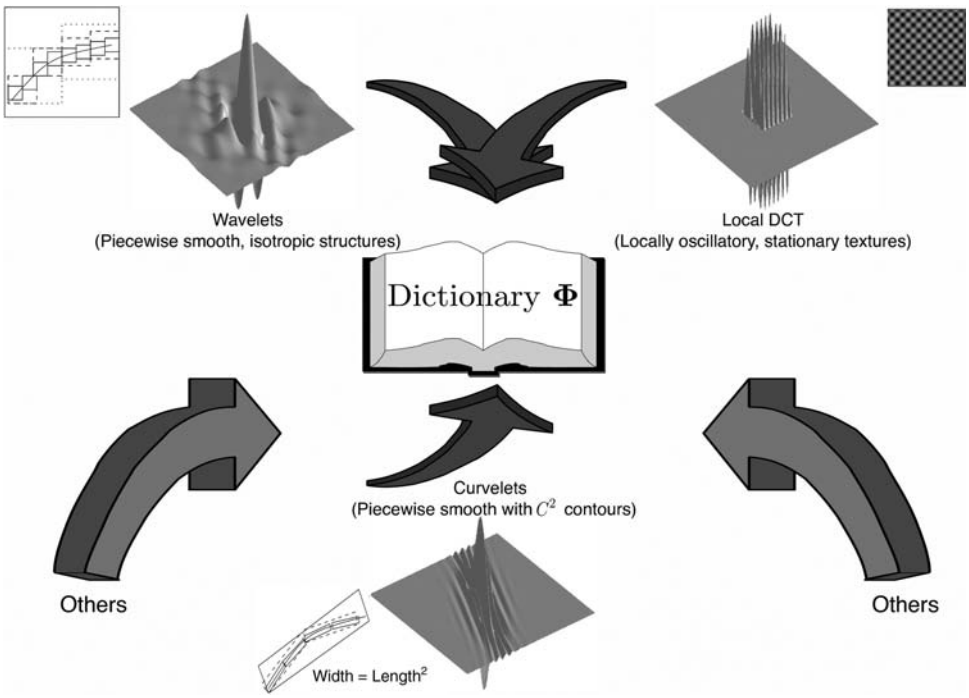

FIGURE 1. Example of a typical dictionary in image processing applications.

link is the identity for the Gaussian case, and log in the Poisson case.

We are seeking a regularized estimate of $\boldsymbol{\alpha}$ from $\mathbf{y}$ using the penalized maximum-likelihood estimator (PMLE)

$$\hat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \ell\ell(\mathbf{y}|\boldsymbol{\alpha}) + \log p_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) \tag{4}$$

which is also termed the maximum *a posteriori* (MAP) estimator. The first term in Equation (4) is the log-likelihood score, which encourages fidelity to the observation. $p_{\boldsymbol{\alpha}}(.)$ stands for the prior imposed on the coefficients $\boldsymbol{\alpha}$, e.g. Gaussian, Laplacian, generalized Gaussian, Gaussian scale-mixture, etc.

Although the above formulation can be abstracted to the exponential family, solving Equation (4) in general is quite a challenging task. We postpone this aspect to future work. We here restrict ourselves to the AWGN case, i.e.

$$\mathbf{Y}|\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\Phi}\boldsymbol{\alpha}, \sigma^2\mathbf{I}). \tag{5}$$

Therefore, the MAP/PMLE estimation problem can then be expressed in terms of the decomposition coefficients $\boldsymbol{\alpha}$, which gives, for AWGN with variance σ^2

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \frac{1}{2\sigma^2}\|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda\Psi(\boldsymbol{\alpha}), \tag{6}$$

where $\lambda > 0$ is the regularization parameter, $\Psi\colon \mathbb{R}^L \mapsto \mathbb{R}^+$ is a penalty function (PF) promoting reconstruction with low complexity taking advantage of sparsity. In the sequel, we additionally assume that the prior associated with $\Psi(\boldsymbol{\alpha})$ is separable, that is, the coefficients $\boldsymbol{\alpha}$ are independent and identically distributed. Hence

$$\Psi(\boldsymbol{\alpha}) = \sum_{l=1}^{L} \psi(\boldsymbol{\alpha}_l). \tag{7}$$

Let us add that Equation (6) is an augmented Lagrangian form of the problem

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \Psi(\boldsymbol{\alpha}) \quad \text{subject to } \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 \leq \tau \tag{8}$$

with an appropriate correspondence between τ and the reciprocal of the Lagrangian multiplier λ.

The regularizing PF ψ is often assumed a non-negative, continuous, even-symmetric and non-decreasing function on $\mathbb{R}^+$, but is not necessarily convex on $\mathbb{R}^+$ and could be irregular at point zero to produce sparse solutions. A popular choice of ψ is the ℓ_1-norm penalty which leads to the well-known soft thresholding rule. The problem in Equation (6) with the ℓ_1-norm penalty is a noise-aware variant of basis pursuit. This has been considered in [50] (BPDN algorithm). Although practical computational complexity of BPDN for large-scale problems (e.g. image processing) is an issue, except when the dictionary is the union of bases, in which case the block coordinate relaxation (BCR) algorithm was

proposed as a fast and effective algorithm [51]. Minimizers of Equation (6) for a general class of regularizing functions (neither non-necessarily differentiable nor convex) have been characterized in a rigorous and systematic way for the orthogonal wavelet transform in [52], and general $\mathbf{\Phi}$ in [53]. However, in the latter, implementation issues were not considered except for the simple orthogonal case. Inspired by the BCR, Elad [54] recently proposed a parallel coordinate descent approach for iteratively solving Equation (6) with overcomplete $\mathbf{\Phi}$ and the ℓ_1-norm penalty. A convergence analysis of this algorithm was recently proposed in [55], but only with a smoothed and strictly convexified version of the ℓ_1 penalty, which is an important and strong assumption.

3.2. Derivation of E and M steps

Let us now turn to the missing data case and let us write $\mathbf{y} = (\mathbf{y}_o, \mathbf{y}_{\text{miss}})$, where $\mathbf{y}_{\text{miss}} = \{\mathbf{y}_i\}_{i \in I_m}$ is the missing data, and $\mathbf{y}_o = \{\mathbf{y}_i\}_{i \in I_o}$. The incomplete observations do not contain all information to apply standard methods to solve Equation (6) and get the PMLE of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^{\mathrm{T}}, \sigma^2)^{\mathrm{T}} \in \Theta \subset \mathbb{R}^L \times \mathbb{R}^{+*}$. Nevertheless, the EM algorithm can be applied to iteratively reconstruct the missing data and then solve Equation (6) for the new estimate. The estimates are iteratively refined until convergence.

3.2.1. The E step

This step computes the conditional expectation of the penalized log-likelihood of complete data, given $\mathbf{y}_o$ and current parameters $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\alpha}^{(t)^{\mathrm{T}}}, \sigma^{2^{(t)}})^{\mathrm{T}}$

$$\begin{aligned} Q\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\big) &= \mathsf{E}\big[\ell\ell(\mathbf{y}|\boldsymbol{\theta}) - \lambda\Psi(\boldsymbol{\alpha})|\mathbf{y}_o, \boldsymbol{\theta}^{(t)}\big] \\ &= \mathsf{E}\big[\ell\ell(\mathbf{y}|\boldsymbol{\theta})|\mathbf{y}_o, \boldsymbol{\theta}^{(t)}\big] - \lambda\Psi(\boldsymbol{\alpha}). \end{aligned} \tag{9}$$

For regular exponential families, the E step reduces to finding the expected values of the sufficient statistics of the complete data $\mathbf{y}$ given observed data $\mathbf{y}_o$ and the estimate of $\boldsymbol{\alpha}^{(t)}$ and $\sigma^{2^{(t)}}$. Then, as the noise is zero-mean white Gaussian, the E-step reduces to calculating the conditional expected values and the conditional expected squared values of the missing data, that is

$$\begin{aligned} \mathbf{y}_i^{(t)} &= \mathsf{E}\left[\mathbf{y}_i|\mathbf{\Phi}, \mathbf{y}_o, \boldsymbol{\alpha}^{(t)}, \sigma^{2^{(t)}}\right] \\ &= \begin{cases} \mathbf{y}_i, & \text{for observed data, } i \in I_o, \\ (\mathbf{\Phi}\boldsymbol{\alpha}^{(t)})_i & \text{for missing data, } i \in I_m, \end{cases} \\ \text{and } \mathsf{E}\left[\mathbf{y}_i^2|\Phi, \mathbf{y}_o, \boldsymbol{\alpha}^{(t)}, \sigma^{2^{(t)}}\right] &= \begin{cases} \mathbf{y}_i^2, & i \in I_o, \\ (\mathbf{\Phi}\boldsymbol{\alpha}^{(t)})_i^2 + \sigma^{2^{(t)}}, & i \in I_m. \end{cases} \end{aligned} \tag{10}$$

3.2.2. The M step

This step consists of maximizing the penalized surrogate function with the missing observations replaced by their estimates in the E step at iteration t, that is

$$\boldsymbol{\theta}^{(t+1)} = \arg\min_{\boldsymbol{\theta} \in \Theta} -Q\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\big). \tag{11}$$

Thus, the M step updates $\sigma^{2^{(t+1)}}$ according to

$$\sigma^{2^{(t+1)}} = \frac{1}{n}\left[\sum_{i \in I_o}\big(\mathbf{y}_i - \big(\mathbf{\Phi}\boldsymbol{\alpha}^{(t)}\big)_i\big)^2 + (n - n_o)\sigma^{2^{(t)}}\right], \tag{12}$$

where $n_o = \text{tr}\mathcal{M} = \text{Card}I_o$ is the number of observed pixels. Note that at convergence, we will have

$$\hat{\sigma}^2 \longrightarrow \frac{1}{n_o}\sum_{i \in I_o}(\mathbf{y}_i - (\mathbf{\Phi}\hat{\boldsymbol{\alpha}})_i)^2$$

which is the maximum-likelihood estimate of the noise variance inside the mask (i.e. with observed pixels).

As far as the update equation of $\boldsymbol{\alpha}^{(t+1)}$ is concerned, this depends not only on the structure of the dictionary $\mathbf{\Phi}$ (i.e. basis, tight frame, etc.), but also on the properties of the PF ψ (i.e. smooth/non-smooth, convex/non-convex). These two parameters also have a clear impact on the convergence behavior of the inpainting algorithm. Important cases will be treated separately in Section 4. In all cases, the overall structure of the EM-based inpainting algorithm is as follows.

Algorithm 1

Require: Observed image $\mathbf{y}_{\text{obs}} = \mathcal{M}\mathbf{y}$ and a mask $\mathcal{M}$, $t = 0$, initial $\boldsymbol{\alpha}^{(0)}$

1: **repeat**
2: ***E Step***
3: Update the image estimate:
$$\mathbf{y}^{(t)} = \mathbf{y}_{\text{obs}} + (\mathbf{I} - \mathcal{M})\mathbf{\Phi}\boldsymbol{\alpha}^{(t)} = \mathbf{\Phi}\boldsymbol{\alpha}^{(t)} + (\mathbf{y}_{\text{obs}} - \mathcal{M}\mathbf{\Phi}\boldsymbol{\alpha}^{(t)}) \tag{13}$$
4: ***M Step***
5: Solve Equation 6 with reconstructed $\mathbf{y}^{(t)}$ substituted for $\mathbf{y}$, and get $\boldsymbol{\alpha}^{(t+1)}$.
6: If desired, update $\sigma^{2^{(t+1)}}$ according to Equation 12.
7: $t = t + 1$,
8: **until** Convergence.
9: Reconstruct $\hat{\mathbf{x}}$ from $\hat{\boldsymbol{\alpha}}$.

If σ^2 happens to be known, step 6 can be dropped from the updating scheme.

4. CONVERGENCE PROPERTIES

In this section, we shall give both the updating rules of $\boldsymbol{\alpha}^{(t+1)}$ and the convergence properties of the corresponding EM

algorithm (and its variants) for different penalties and types of dictionaries (the columns of which are always supposed to be normalized to a unit ℓ_2-norm). Convergence is examined in both the weak and strong topologies.

The convergence properties of the EM algorithm and its generalizations [e.g. GEM, expectation conditional maximization (ECM), see [38, 56]] have deeply been investigated by many authors (see, e.g. [37, 38, 57]). However, most of their results heavily rely on the smoothness of the log-likelihood function or its penalized form, hence on the regularity of the penalty ψ (e.g. $Q(.|.)$ is C^1 for many results to hold [57]). But, to produce sparse solutions, we are mainly concerned with non-smooth penalties. Thus, most of classical convergence results of the EM and its variants do not hold here. We therefore need more involved arguments that we mainly borrow from non-smooth convex analysis and the optimization literature.

4.1. $\boldsymbol{\Phi}$ is a basis

In this case, $\boldsymbol{\Phi}$ is a bijective bounded linear operator. Algorithm 1 is an EM, which can be solved for the coefficients $\boldsymbol{\alpha}$, or equivalently, in terms of the image samples $\mathbf{x}$.

Many sparsity promoting PFs have been proposed in the literature. Although they differ in convexity, boundedness, smoothness and differentiability, they generally share some common features that we state here as general conditions for a family of PFs considered in our setting. Suppose that:

- H1. ψ is even-symmetric, non-negative and non-decreasing on $[0, +\infty)$, and $\psi(0) = 0$.
- H2. ψ is twice differentiable on $\mathbb{R}\backslash\{0\}$, but not necessarily convex.
- H3. ψ is continuous on $\mathbb{R}$, it is not necessarily smooth at zero and admits a positive right derivative at zero $\psi'_+(0) = \lim_{h\to 0^+} (\psi(h)/h) > 0$.
- H4. The function $\alpha + \lambda\psi'(\alpha)$ is unimodal on $(0, +\infty)$.

Among the most popular PFs satisfying the above requirements, we cite $\psi(\alpha) = |\alpha|^p$, $0 < p \leq 2$ (generalized Gaussian prior), $\psi(\alpha) = \sqrt{(\alpha^2 + \nu)}$, $\nu \geq 0$, the SCAD penalties [52]; see, e.g. [52, 58] for a review of others.

PROPOSITION 1. *If ψ respects H1–H4,*

(i) *The step has exactly one solution decoupled in each coordinate $z_l = \phi_l^{\mathrm{T}}\, \mathbf{y}^{(t)}$:*

$$\hat{\boldsymbol{\alpha}}_l(z_l) = \begin{cases} 0, & \text{if} |z_l| \leq \lambda\sigma^2\psi'_+(0), \\ z_l - \lambda\sigma^2\psi'(\hat{\boldsymbol{\alpha}}_l), & \text{if} |z_l| > \lambda\sigma^2\psi'_+(0), \end{cases} \quad (14)$$

which is continuous in z_l.

(ii) *The sequence of observed penalized likelihoods converges (actually increases) monotonically to a stationary point of the penalized likelihood.*

(iii) *All limit points of the EM inpainting sequence $\{\mathbf{x}^{(t)}, t \geq 0\}$ generated by Algorithm 1 are stationary points of the penalized likelihood.*

See Appendix for the proof.

PROPOSITION 2. *If ψ satisfies H1–H4, is coercive and proper convex.*

(i) *The inpainting problem has at least one solution.*

(ii) *The inpainting problem has exactly one solution if ψ is strictly convex.*

(iii) *The instance of iterates $\{\mathbf{x}^{(t)}, t \geq 0\}$ is asymptotically regular, i.e. $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\| \to 0$, and converges weakly to a fixed point.*

(iv) *If $\psi(\alpha) = |\alpha|^p$, $1 \leq p \leq 2$, the sequence of iterates converges strongly to a solution of the inpainting problem.*

See Appendix for the proof.

REMARK 1.

(i) When $\psi(\alpha) = |\alpha|$, the shrinkage operator is the popular soft thresholding rule. The EM inpainting with a basis is then an iterative soft thresholding, which is strongly convergent. Strong convergence can also happen with other penalties than $\psi(\alpha) = |\alpha|^p, p \in [1, 2]$. For instance, any strongly convex function ψ will imply strong convergence (see [59, Corollary 5.16 and Proposition 3.6]).

(ii) The sequence of inpainting iterates is asymptotically regular. This furnishes a simple stopping rule test, typically $\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\| < \delta$ for sufficiently small $\delta > 0$.

(iii) If one has additional constraints to be imposed on the solution such as forcing the image to lie in some closed convex subset of $\mathcal{H}$ (e.g. positive cone), the iteration in Equation (A.2) can be easily adapted by incorporating the projector onto the constraint set, e.g. replace $\mathcal{D}_{\boldsymbol{\Phi},\lambda}$ by $\mathbf{P}_+ \circ \mathcal{D}_{\Phi,\lambda}$, $(\mathbf{P}_+\mathbf{x})_i = \max(\mathbf{x}_i, 0)$.

4.2. $\boldsymbol{\Phi}$ is a frame

$\boldsymbol{\Phi}$ is no longer a bijective linear operator (surjective but not injective). Consequently, the EM step will be solved in terms of the transform coefficients. The inpainted image is reconstructed after convergence of the EM algorithm. This is a major distinction with the previous case of a basis. Indeed, when solving the problem in terms of image (or signal) samples, the solution $\mathbf{x}$ is allowed to be an arbitrary vector over $\mathbb{R}^n$, whereas the solution image is confined to the column space of $\boldsymbol{\Phi}$ when solving in terms of coefficients $\boldsymbol{\alpha}$. This distinction holds true except when $\boldsymbol{\Phi}$ is orthogonal. We therefore propose the following proximal Landweber-like iterative scheme, which summarizes the E and M steps

$$\boldsymbol{\alpha}^{(t+1)} = S_{\mu\lambda\sigma^2}\big(\boldsymbol{\alpha}^{(t)} + \mu\boldsymbol{\Phi}^{\mathrm{T}}\big(\mathbf{y}_{\mathrm{obs}} - \mathcal{M}\boldsymbol{\Phi}\boldsymbol{\alpha}^{(t)}\big)\big), \quad \mu > 0, \quad (15)$$

where $S_{\mu\lambda\sigma^2}$ is the componentwise shrinkage operator associated with Equation (14) with parameter $\mu\lambda\sigma^2$. Again, the questions we want to answer are, for a general frame, (i) does this algorithm converge, (ii) what is the convergence behavior of such an iterative scheme, (iii) can the iteration be computed efficiently? The answer to the first question is yes under assumptions similar to the ones stated for the basis case. This will be summarized shortly in a general result together with an answer to question (ii). As far as point (iii) is concerned, a positive answer cannot be given in general since multiplications by $\mathbf{\Phi}$ and $\mathbf{\Phi}^{\mathrm{T}}$ are involved. A very interesting case happens when the dictionary is a tight frame. In such a case, following the discussion in Section 2 (see also Remark 2), multiplications by $\mathbf{\Phi}$ and $\mathbf{T}$ are never explicitly implemented (fast implicit analysis and synthesis operators are used instead).

We are now ready to state a convergence result for frames

PROPOSITION 3. *Suppose that ψ satisfies H1–H4, is coercive and proper convex. Moreover, suppose that $0 < \mu < 2/B$, where B is the frame upper bound. Then the following holds*:

(i) *The inpainting problem has at least one solution.*
(ii) *The inpainting problem has exactly one solution if ψ is strictly convex.*
(iii) *The instance of coefficient iterates $\{\boldsymbol{\alpha}^{(t)},\ t \geq 0\}$ is asymptotically regular and converges weakly to a solution of the inpainting algorithm.*
(iv) *If $\psi(\alpha) = |\alpha|^p$, $1 \leq p \leq 2$, the sequence of coefficient iterates converges strongly to a solution of the inpainting problem.*

See Appendix for the proof.

REMARK 2.

(i) Again, for $\psi(\alpha) = |\alpha|$, the EM inpainting with a frame is a strongly convergent iterative soft thresholding. Note also that the strong convexity of ψ is sufficient for assertion (iv) in Proposition 3 to remain true.
(ii) Note that with frames, it becomes less straightforward to impose some side constraints on the solution such as positivity, compared to the previous case of bases.
(iii) An interesting special case of Equation (15) arises when $\mathbf{\Phi}$ is a tight frame ($A = B$, $\mathbf{T}^{+} = B^{-1}\mathbf{\Phi}$), $\mu = B^{-1}$ [2], and $\psi(\alpha) = |\alpha|$. Letting $\boldsymbol{\kappa}^{(t)} = \mathbf{T}\mathbf{y}_{\mathrm{obs}}/B + (\mathbf{I} - \mathbf{T}\mathcal{M}\mathbf{T}^{+})\boldsymbol{\alpha}^{(t)}$, and specializing Equation (14) to soft thresholding, Equation (15) becomes

$$\boldsymbol{\alpha}_l^{(t+1)} = \begin{cases} 0, & \text{if } |\boldsymbol{\kappa}_l^{(t)}| \leq \lambda\sigma^2/B, \\ \boldsymbol{\kappa}_l^{(t)} - \mathrm{sign}(\boldsymbol{\kappa}_l^{(t)})\lambda\sigma^2/B, & \text{if } |\boldsymbol{\kappa}_l^{(t)}| > \lambda\sigma^2/B. \end{cases} \tag{16}$$

Except for the scaling by B, it turns out that this iteration is a simple shrinkage applied to the data reconstructed at the E step. It embodies an (implicit) reconstruction step followed by an analysis and a shrinkage step, and is very efficient to compute. A similar iteration can be developed for any other rule corresponding to any other appropriate penalty.

4.3. $\mathbf{\Phi}$ is a union of transforms

Let us now assume the general case where the dictionary is a union of K sufficiently incoherent transforms, that is $\mathbf{\Phi} = [\mathbf{\Phi_1}, \ldots, \mathbf{\Phi_K}]$. In this case, the M step of the inpainting becomes more complicated and computationally unattractive. However, this maximization becomes much simpler when carried out with respect to each transform coefficient with the other coefficients held constant. In the statistical literature, this generalized version of the EM is termed ECM [56, 60, 61]. Without the E step, this algorithm is also known as (block) coordinate relaxation in optimization theory [62, 63].

More precisely, ECM replaces each M step of the EM, by a cyclic sequence of K conditional optimization steps, each of which minimizes the surrogate functional Q over $\boldsymbol{\alpha_k}$, $k \in \{1, \ldots, K\}$, with the other transform coefficients $\boldsymbol{\alpha}_{k' \neq k}$ fixed at their previous value. The M step in Algorithm 1 then becomes (we omit the update of σ for briefness),

---

Compute the residual $\mathbf{r}^{(t)} = \mathbf{y}_{\mathrm{obs}} - \mathcal{M}\sum_{k=1}^{K} \mathbf{\Phi_k}\boldsymbol{\alpha}_k^{(t)}$,
for Each transform $k \in \{1, \ldots, K\}$ in the dictionary **do**
CM Step: Solve the optimization problem

$$\boldsymbol{\alpha}_k^{(t+1)} = \arg\min_{\boldsymbol{\alpha_k}} \frac{1}{2\sigma^2} \left\| \left(\mathbf{r}^{(t)} + \mathbf{\Phi_k}\boldsymbol{\alpha}_k^{(t)}\right) - \mathbf{\Phi_k}\boldsymbol{\alpha_k} \right\|_2^2 + \lambda_k \Psi(\boldsymbol{\alpha_k}) \tag{17}$$

end for

The optimization subproblem for each k can be solved by adapting results of the previous two sections depending on whether $\mathbf{\Phi_k}$ is a basis or a tight frame. Let us also mention that a variation of the above ECM, termed multi-cycle ECM in [56], would be to include an E step before each CM step, which is equivalent to putting the E and CM steps inside the loop over transforms.

Under appropriate regularity conditions on the complete data penalized likelihood, it has been shown that the ECM extension (which is actually a Generalized EM; see [37]) shares all the appealing convergence properties of the GEM such as monotonically increasing the likelihood, convergence (in the weak topology) to a stationary point, local or global minimum, etc. (see [56, 57] for more details). However, for many of their results to apply, smoothness and strict convexity of the PF are again important assumptions (PF must be C^1),

[2] See the proof of Proposition 3 for the definition of A and B.

which is not the case for the sparsity-promoting penalties adopted here.

One strategy to circumvent this difficulty is to smooth at zero the family of PFs we consider. Typically, many PFs fulfilling requirements H1–H4 may be recast as functions of $|\alpha|$. Then, one may operate the change of variable in $\psi(|\alpha|)$ to $\psi(\sqrt{(\alpha^2+\nu)})$ with sufficiently small $\nu > 0$. For example, the PF corresponding to the ℓ_p-norm may be replaced with $(\alpha^2+\nu)^{p/2}$, or with the function proposed in [55]. The latter approximation has the advantage of providing a closed-form, but approximate solution to Equation (14). Note also that these approximations are strict convexification of the original ℓ_p PF. Hence, with the latter smooth and strictly convex PF approximations, by straightforward adaptation of the results in [56, 57], it can easily be shown that the penalized likelihood converges monotonically to a stationary point, and that the sequence of multi-cycle ECM iterates converges (weakly) to the unique stationary point which is also the global maximizer of the penalized likelihood.

The other strategy is to extend convergence results of the ECM to non-smooth PFs. As in the previous sections, this will be accomplished by exploiting the separable structure of PF and appealing to more elaborated tools in optimization theory, namely (block) coordinate relaxation-based minimization methods for non-differentiable functions [63, 64]. Note that the BCR algorithm has already been used for ℓ_1- regularized sparse denoising problems with overcomplete dictionaries, but only for dictionaries which are a union of bases, while our algorithm stated below will be valid even for more complicated dictionaries and PFs.

4.3.1. *General algorithm and its properties*

Here is now our general multi-cycle ECM inpainting algorithm for dictionaries that are a union of several (incoherent) transforms:

Algorithm 2

Require: Observed image $\mathbf{y}_{\text{obs}}$ and a mask $\mathcal{M}$, $t=0$, initial $\boldsymbol{\alpha}^{(0)}$, $\forall k \in \{1,\dots,K\}$, $0 < \mu_k < 2/B_k$,

1: **repeat**
2: Select a transform $k \in \{1,\dots,K\}$ with a cyclic rule.
3: $\mathbf{r}^{(t)} = \mathbf{y}_{\text{obs}} - \mathcal{M}\sum_{k=1}^{K} \boldsymbol{\Phi}_k \boldsymbol{\alpha}_k^{(t)}$,
4: **for** $m = 1$ to M **do**
5: Update,

$$\boldsymbol{\alpha}_k^{(t+m/M)} = S_{\mu_k\lambda_k\sigma^2}\Big[\boldsymbol{\alpha}_k^{(t+(m-1)/M)} + \mu_k \boldsymbol{\Phi}_k^T\Big(\Big(\boldsymbol{r}^{(t)} + \mathcal{M}\boldsymbol{\Phi}_k\boldsymbol{\alpha}_k^{(t)}\Big) - \mathcal{M}\boldsymbol{\Phi}_k\boldsymbol{\alpha}_k^{(t+(m-1)/M)}\Big)\Big] \quad (18)$$

6: **end for**
7: $t = t+1$,
8: **until** Convergence.
9: Reconstruct $\hat{\mathbf{x}}$ from $\hat{\boldsymbol{\alpha}}$.

Note that for a single transform, this algorithm reduces to Equations (A.2) and (15) for respective corresponding cases.

The following result establishes the convergence behavior of the multi-cycle ECM.

PROPOSITION 4 *Suppose that ψ is a coercive proper convex satisfying H1–H4. Then, the penalized likelihood converges (increasingly) monotonically, and the sequence of coefficient iterates $\{\boldsymbol{\alpha}^{(t)}, t \geq 0\}$ generated by the multi-cycle ECM Algorithm 2 is defined, bounded and converges to a solution of the inpainting algorithm.*

See Appendix for the proof.

4.4. Choice of the regularization parameter

In this section, we aim at giving some guidelines for alleviating the somewhat delicate and challenging problem of choosing the regularization parameter λ. So far, we have characterized the solution for a particular choice of λ. In particular, for sufficiently large λ (e.g. $\lambda \geq \|\mathbf{T}\mathbf{y}_{\text{obs}}\|_\infty/\sigma^2$ for the ℓ_1-norm PF), the inpainting solution will coincide with the sparsest one (i.e. $\boldsymbol{\alpha} = 0$). But this will be at the price of very bad data fidelity, and thus λ is a regularization parameter that should not be chosen too large.

One attractive solution is based upon the following observation. At early stages of the algorithm, posterior distribution of $\boldsymbol{\alpha}$ is unreliable because of missing data. One should then consider a large value of λ (e.g. $+\infty$ or equivalently $\lambda \geq \|\mathbf{T}\mathbf{y}_{\text{obs}}\|_\infty/\sigma^2$ for the ℓ_1-norm PF) to favor the penalty term. Then, λ is monotonically decreased (e.g. according to a linear schedule). For each $\lambda\sigma^2 \geq \tau\sigma$ (where τ is typically 3 to 4 to reject the noise), the ECM inpainting solution is computed and serves as an initialization for the next value of λ. This procedure has a flavor of deterministic annealing [65], where the role of the regularization parameter parallels that of the temperature. This procedure is also closely related to homotopy continuation and path following methods [66] and [67]. These aspects are now under investigation and will be deferred to a future paper.

5. EXPERIMENTAL RESULTS

To illustrate the potential applicability of the ECM inpainting algorithm, we applied it to a wide range of problems including inpainting, interpolation of missing data and zooming. Both synthetic and real images are tested. In all experiments, the convex ℓ_1 penalty was used. For the case of dictionaries built from multiple transforms, and with the multi-cycle ECM Algorithm 2, λ was the same for all transforms and $\mu_k = B_k^{-1}$, $\forall k \in \{1,\dots,K\}$.

The first example is depicted in Fig. 2 with Lena, where 80% pixels were missing, with huge gaps. The dictionary contained the fast curvelet transform (Second generation tight frame curvelets [42]). The final threshold parameter was fixed to the value 3σ. This example is very challenging, and the inpainting algorithm performed impressively well. It managed to recover most important details of the image that

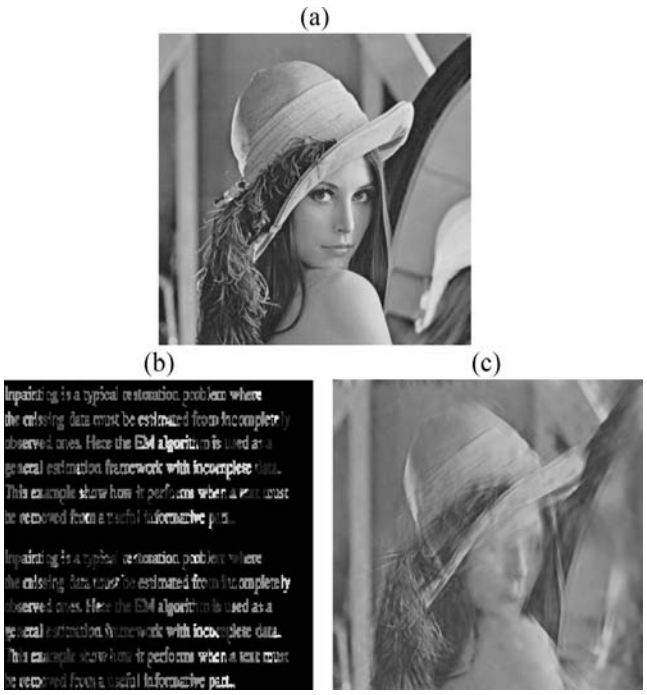

FIGURE 2. Example with Lena. Dictionary: curvelets, penalty: ℓ_1. (a) Original. (b) Masked with 80% pixel missing (missing pixels are in black) $\text{PSNR}_{\text{degraded}} = 6.43$ dB. (c) Inpainted $\text{PSNR}_{\text{restored}} = 19.7$ dB.

are almost impossible to distinguish in the masked image. The visual quality is confirmed by values of the peak signal-to-noise ratio (PSNR) as reported in Fig. 2.

The algorithm was also applied to a superresolution problem (see Fig. 3). In this example, the image of Claudia was regularly downsampled with 50% missing 8×8 pixel blocks (Fig. 3b). The dictionary again included the fast curvelet transform. The result is shown in Fig. 3c). The image main features and fine details were recovered with a little distortion, which is also confirmed by the measures of the PSNR.

To further illustrate the power of the ECM inpainting algorithm, we applied it to the challenging Barbara textured image. As stationary textures are efficiently represented by the local DCT, the dictionary contained both the fast curvelet (for the geometry part) and the LDCT transforms. In this experiment, the Barbara image was masked with an increasing number of missing pixels in the range 20–80%. The results for 50 and 80% missing pixels are portrayed in Fig. 4. Despite a large amount of missing pixels, the algorithm is not only able to recover the geometric part (cartoon), but particularly performs well inside the difficult textured areas, e.g. trousers (it compares very favorably to [22, Fig. 2] and [68, Figs. 14–16]).

The algorithm was finally tested on two real photographs of a parrot and a zebra,[3] where we wanted to remove the grid of the cage to virtually free the bird and zebra (i.e. disocclusion example). The mask of the grid was manually plotted. For the color image, each of the three color bands (in RGB space) was processed independently and then re-merged in the final color result image. The results are shown in Figs. 5 and 6. The dictionary contained the undecimated DWT and LDCT for the parrot and the fast curvelet transform for the zebra. For comparative purposes, inpainting results of a PDE-based method [5] are also reported in each case. For the zebra example, the curvilinear structures of the black stripes are better recovered with ECM than with the PDE-based method (see, e.g. the neck and the back of zebra). In large smooth regions, both methods are quite equivalent, and neither outperforms the other. As far as the parrot example is concerned, as expected, the main differences between the two approaches are essentially located in the 'textured' area in the vicinity of the parrot's eye. Indeed, pure diffusion or transport PDE-based methods have limited performance when dealing with textures. Nonetheless, one must avoid inferring stronger statements (such as which is better than which in general), as our comparison was carried out on a small sample of images with one PDE-based inpainting method. Further comparative work is needed to answer these questions more precisely.

5.1. Reproducible research

Following the philosophy of reproducible research [69], the MCALab toolbox is made available freely for download at http://www.greyc.ensicaen.fr/~jfadili/demos/WaveRestore/downloads/mcalab.html. MCALab is a collection of Matlab functions, scripts and graphical user interface for MCA [31] and ECM inpainting of 1-D signals and 2-D images. MCALab requires at least WaveLab 8.02 [69] to work properly. MCALab has been used by the authors to produce some of the figures included in this paper. A C++ library has also been designed and will be available on the web very soon.

6. RELATED WORK

In the same vein as [29], the authors [68, 70, 71] have also formulated the inpainting problem using sparse decompositions. Although these works share similarities with our method, they differ in many important aspects. For instance, in [68, 70], the inpainting problem is stated as a denoising task in two steps: estimation of the image assuming the support is known (i.e. set of non-null representation coefficients), and a progressive estimation of the support. All the theoretical analysis is carried out assuming the support is known, which is obviously not the case with real data. In our work, the inpainting problem is stated as a linear inverse ill-posed problem, that is solved in a principled Bayesian framework, and for which the EM mechanism comes as a

[3] Further examples and color images may be found at the http://www.greyc.ensicaen.fr/~jfadili/themes.html and http://jstarck.free.fr/mca.html.

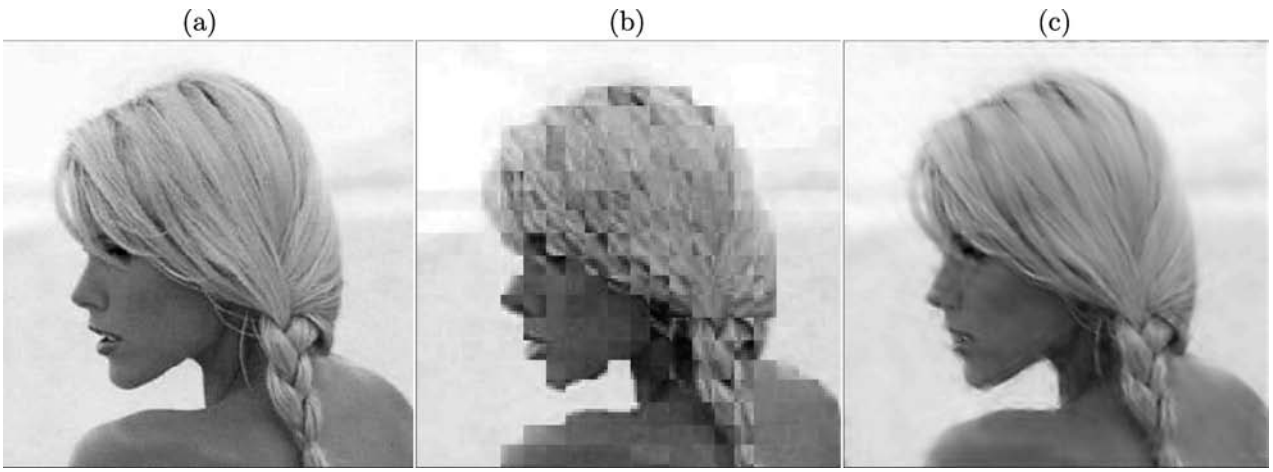

FIGURE 3. Application to superresolution (50% missing 8×8 pixel blocks). Dictionary: curvelets, penalty: ℓ_1. (a) Original image. (b) Low-resolution image $\text{PSNR}_{\text{degraded}} = 5.52$ dB. (c) Interpolated image $\text{PSNR}_{\text{restored}} = 30.82$ dB.

FIGURE 4. Example with Barbara. Dictionary: curvelets + LDCT, penalty: ℓ_1. (a) Original image. (b) and (c) Masked images with 50% ($\text{PSNR}_{\text{degraded}} = 8.6$ dB) and 80% ($\text{PSNR}_{\text{degraded}} = 6.54$ dB) missing pixels. (d) and (e) and Inpainted images corresponding to (b) and (c). (d) $\text{PSNR}_{\text{restored}} = 33.4$ dB. (e) $\text{PSNR}_{\text{restored}} = 27.4$ dB.

natural iterative algorithm because of physically missing data. Note also that our framework is more general as it allows estimation of the hyperparameters associated with the noise model as a by-product, not to mention that it can be generalized to handle other noises and other inverse problems than inpainting as we explained in Section 3.1. Furthermore, in our work, we give a detailed and theoretically grounded convergence analysis to characterize the type of solutions. Such a

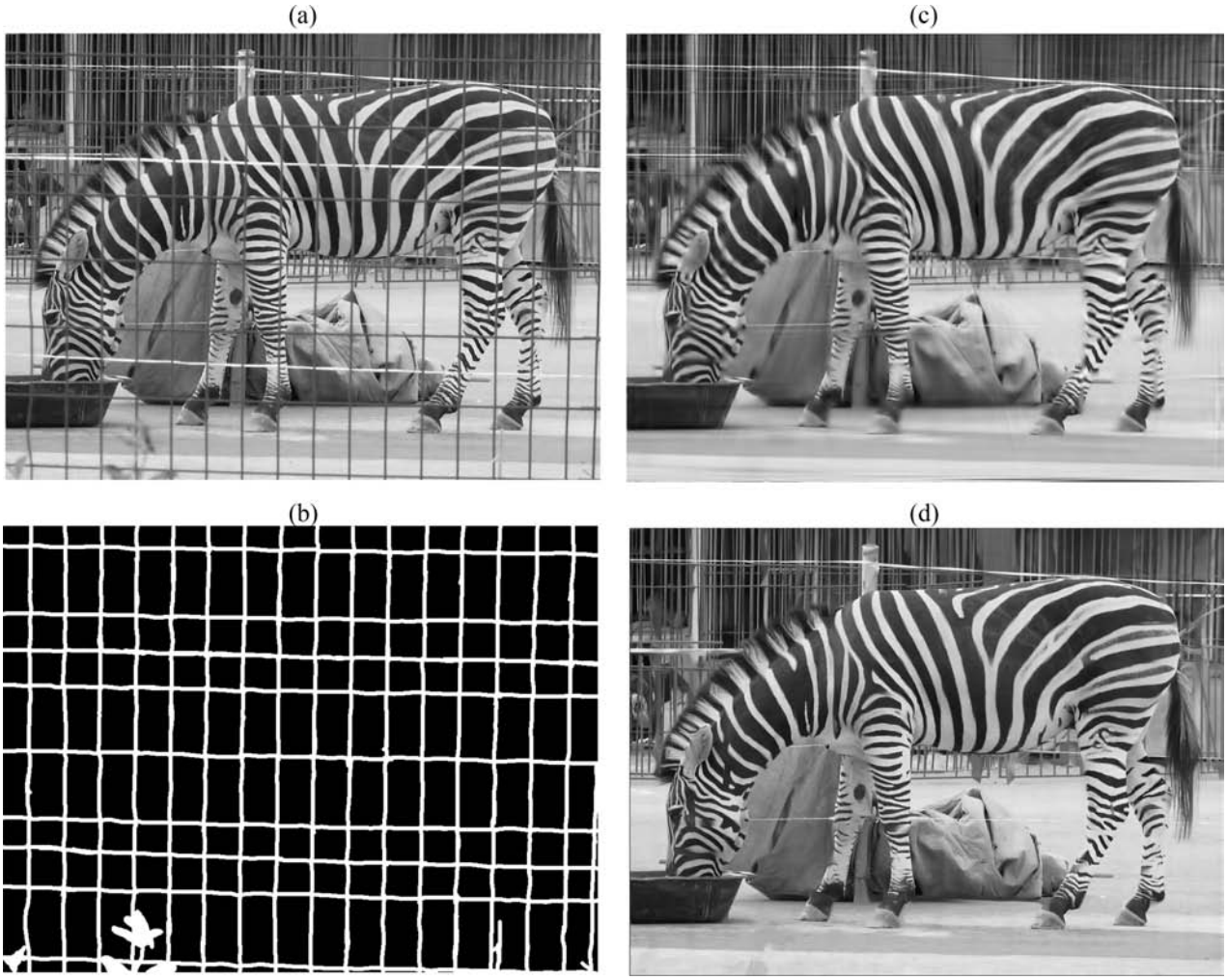

FIGURE 5. Free the zebra by removing the cage. (a) Original. (b) Mask (missing pixels are in white). (c) ECM inpainting (dictionary: curvelets). (d) PDE inpainting.

convergence study is lacking in the work of [70, 71]. The method in [70] can only handle dictionaries, which are union of orthonormal bases, and the one in [71] proposes the same algorithm as [70] with a dictionary formed by a single tight frame (framelet). Our method seems to be more general and naturally deals with any dictionary (union of transforms that correspond to frames or tight frames such as the curvelet transform). Hence, our algorithm can be used with richer overcomplete dictionaries (beyond union of bases) that are able to provide us with sparser representations of a wide range of images based on their morphological content.

Additionally, we would like to make a connection with the emerging theory which goes by the name of 'compressive sampling' or 'compressed sensing' as recently developed by Donoho [32] and Donoho and Tsiag [33] and Candés *et al.* [34–36]. Roughly speaking, this theory establishes in a principled way that it is possible to reconstruct images or signals accurately from a number of samples which is far smaller than the number of pixels in the image, if the image or signal is sparse (compressible) enough in some dictionary. In fact, accurate and sometimes exact recovery is possible by solving a simple convex optimization problem. Such a theory, although directed toward random sensing matrices, provides a theoretical justification as to why algorithms such as ours work so well in an application such as image inpainting and zooming.

7. SUMMARY AND CONCLUSION

A novel and flexible inpainting algorithm has been presented. Its theoretical properties were systematically investigated and established in various scenarios. To demonstrate potential applications of the algorithm, examples including interpolation, zooming (superresolution) and disocclusion were considered. Several interesting perspectives of this work are under investigation. We can cite the formal investigation of the influence of the regularization parameter (path following/homotopy continuation). Extension to multi-valued images

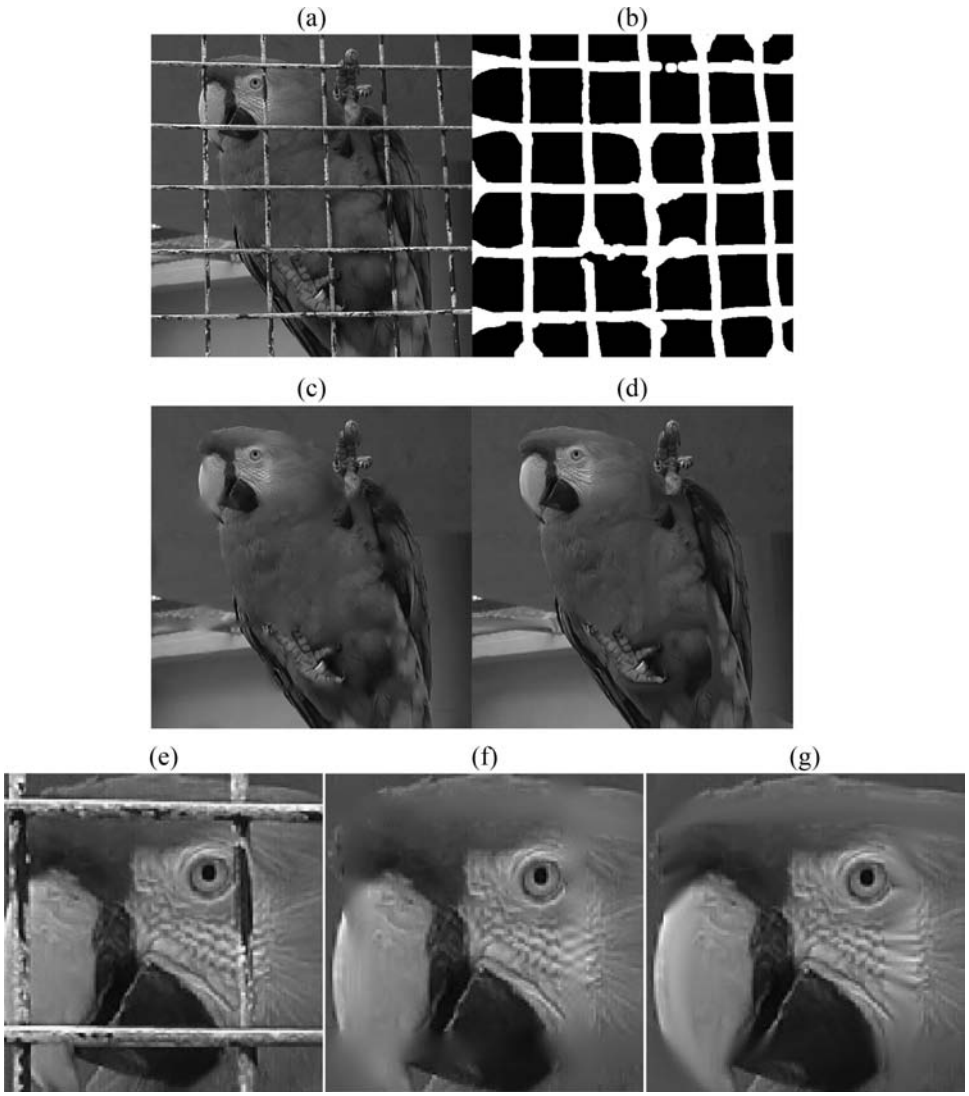

FIGURE 6. Free the parrot by removing the cage. (a) Original. (b) Mask (missing pixels are in white). (c) ECM inpainting (dictionary: wavelets + LDCT). (d) PDE inpainting. (e)–(g) Zoom of (a), (c) and (d).

(e.g. hyper-spectral data) is also an important aspect that is the focus of our current research.

ACKNOWLEDGEMENTS

We would like to thank deeply Miki Elad and Dave Donoho for very interesting and stimulating discussions. We also thank David Tschumperlé for kindly providing his results on image inpainting shown in this paper.

REFERENCES

[1] Masnou, S. and Morel, J. (1998) Level Lines-Based Disocclusion. *IEEE Int. Conf. Image Processing*, Chicago, vol. 3, pp. 259–263.

[2] Masnou, S. and Morel, J. (2002) Disocclusion: a variational approach using level lines. *IEEE Trans. Image Process*, **11**, 68–76.

[3] Bertalmío, M., Sapiro, G., Caselles, V. and Ballester, C. (2000) Image inpainting. *SIGGRAPH 2000*, New Orleans, USA, pp. 417–424.

[4] Bertalmío, M., Bertozzi, A. and Sapiro, G. (2001) Navier–Stokes, fluid-dynamics and image and video inpainting. *CVPR'01*, Kauai, HI, vol. 1, pp. 355–362.

[5] Tschumperlé, D. and Deriche, R. (2005) Vector-valued image regularization with PDEs : a common framework for different applications. *IEEE Trans. Pattern Anal. Mach. Intell.*, **27**, 506–517.

[6] Bornemann, F. and März, T. (2006) Fast image inpainting based on coherence transport. Technical report. Technische Universität München.

[7] Caselles, V., Morel, J. and Sbert, C. (1998) An axiomatic approch to image interpolation. *IEEE Trans. Image Process*, **7**, 376–386.

[8] Ballester, C., Caselles, V., Verdera, J., Sapiro, G. and Bertalmío, M. (2001) A Variational Model for Filling-in Gray Level and Color Images. *IEEE Int. Conf. Computer Vision*, Vancouver, Canada, pp. 10–16.

[9] Ballester, C., Bertalmío, M., Caselles, V., Sapiro, G. and Verdera, J. (2001) Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. Image Process*, **10**, 1200–1211.

[10] Ballester, C., Caselles, V. and Verdera, J. (2003) Disocclusion by joint interpolation of vector fields and gray levels. *Multiscale Modeling and Simul*., **2**, 80–123.

[11] Chan, T. and Shen, J. (2001) Non-texture inpainting by curvature-driven diffusions (cdd). *J. Vis. Commun. Image Represent*., **12**, 436–449.

[12] Chan, T. and Shen, J. (2001) Mathematical models for local non-texture inpainting. *SIAM J. Appl. Math.*, **62**, 1019–1043.

[13] Chan, T., Kang, S.H. and Shen, J. (2002) Euler's elastica and curvature based inpainting. *SIAM J. Appl. Math.*, **63**, 564–592.

[14] Esedoglu, S. and Shen, J. (2002) Digital image inpainting by the Mumford-Shah-Euler image model. *Eur. J. Appl. Math.*, **13**, 353–370.

[15] Chan, T.F., Yip, A.M. and Park, F.E. (2004) Simultaneous total variation image inpainting and blind deconvolution. *Int. J. Imaging Syst. Technol.*, **15**, 92–102.

[16] Chan, T.F., Ng, M.K., Yau, A.C. and Yip, A.M. (2006) Superresolution image reconstruction using fast inpainting algorithms. Technical report. UCLA CAM.

[17] Song, B. (2003) Topics in Variational PDE Image Segmentation, Inpainting and Denoising. PhD Thesis, University of California, LA.

[18] Bertozzi, A., Esedoglu, S. and Gillette, A. (2007) Inpainting by the Cahn-Hilliard equation. *IEEE Trans. Image Process*, **16**, 285–291.

[19] Chan, T.F. and Shen, J.J. (2005) *Image Processing and Analysis-Variational, PDE, Wavelet, and Stochastic Methods*. SIAM, Philadelphia.

[20] Bertalmío, M., Caselles, V., Haro, G. and Sapiro, G. (2005) The State of the Art of Image and Surface Inpainting. In Paragios, N., Faugeras, O. and Chen, Y. (eds.), *Mathematical Models in Computer Vision: The Handbook*, pp. 309–322. Springer-Verlag, Berlin.

[21] Efros, A.A. and Leung, T.K. (1999) Texture Synthesis by Non-Parametric Sampling. *IEEE Int. Conf. Computer Vision*, Kerkyra, Greece, pp. 1033–1038.

[22] Bertalmío, M., Vese, L., Sapiro, G. and Osher, S. (2003) Simultaneous structure and texture image inpainting. *IEEE Trans. Image Process*, **12**, 882–889.

[23] Demanet, L., Song, B. and Chan, T. (2003) Image Inpainting by Correspondence Maps: A Deterministic Approach. *VLSM*, Nice, France.

[24] Rane, S., Bertalmio, M. and Sapiro, G. (2002) Structure and texture filling-in of missing image blocks for wireless transmission and compression applications. *IEEE Trans. Image Process*, **12**, 296–303.

[25] Jia, J. and Tang, C.-K. (2003) Image Repairing: Robust Image Synthesis By Adaptive Nd Tensor Voting. *IEEE CVPR*, vol. 1, pp., 643–650.

[26] Acton, S.T., Mukherjee, D.P., Havlicek, J.P. and Bovik, A.C. (2001) Oriented texture completion by AM-FM reaction-diffusion. *IEEE Trans. Image Process*, **10**, 885–896.

[27] Criminisi, A., Pérez, P. and Toyama, K. (2004) Region filling and object removal by examplar-based image inpainting. *IEEE Trans. Image Process*, **13**, 1200–1212.

[28] Drori, I., Cohen-Or, D. and Yeshurun, H. (2003) Fragment-based image completion. *ACM Trans. Graph.*, **22**, 303–312.

[29] Elad, M., Starck, J.-L., Querre, P. and Donoho, D. (2005) Simultaneous cartoon and texture image inpainting. *Appl. Comput. Harmon. Anal.*, **19**, 340–358.

[30] Starck, J.-L., Elad, M. and Donoho, D. (2005) Image decomposition via the combination of sparse representatntions and variational approach. *IEEE Trans. Image Process*, **14**, 1570–1582.

[31] Starck, J.-L., Elad, M. and Donoho, D. (2004) Redundant multiscale transforms and their application for morphological component analysis. *Adv. Imag. Electron Phys.*, **132**, 287–348.

[32] Donoho, D.L. (2004) Compressed sensing. Technical report. Stanford University, Department of Statistics.

[33] Donoho, D. and Tsaig, Y. (2006) Extensions of compressed sensing. *EURASIP J. Appl. Signal Process.*, **86**, 549–571.

[34] Candès, E., Romberg, J. and Tao, T. (2004) Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. Technical report. CalTech, Applied and Computational Mathematics.

[35] Candès, E. and Tao, T. (2005) Stable signal recovery from noisy and incomplete observations. Technical report. CalTech, Applied and Computational Mathematics.

[36] Candès, E. and Tao, T. (2004) Near optimal signal recovery from random projections: universal encoding strategies? Technical report. CalTech, Applied and Computational Mathematics.

[37] Dempster, A., Laird, N. and Rubin, D. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, **39**, 1–38.

[38] Little, R. and Rubin, D. (1987) *Statistical Analysis with Missing Data*. Wiley, New York.

[39] Mallat, S.G. (1998) *A Wavelet Tour of Signal Processing* (2nd edn). Academic Press.

[40] Starck, J.-L., Murtagh, F. and Bijaoui, A. (1998) *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press.

[41] Candès, E.J. and Donoho, D.L. (1999) Curvelets—a surprisingly effective nonadaptive representation for objects with edges. In Cohen, A., Rabut, C. and Schumaker, L. (eds.), *Curve and Surface Fitting: Saint-Malo 1999*, Vanderbilt University Press, Nashville, TN.

[42] Candès, E., Demanet, L., Donoho, D. and Ying, L. (2006) Fast discrete curvelet transforms. *SIAM Multiscale Modeling Simul.*, **5**, 861–899.

[43] Candès, E. and Donoho, D. (1999) Ridgelets: the key to high dimensional intermittency? *Philos. Trans. R. Soc. Lond. A*, **357**, 2495–2509.

[44] Starck, J.L., Candès, E. and Donoho, D.L. (2002) The curvelet transform for image denoising. *IEEE Trans. Image Process*, **11**, 670–684.

[45] Donoho, D. and Elad, M. (2003) Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization. *Proc. Natl Acad. Sci.*, **100**, 2197–2202.

[46] Bruckstein, A. and Elad, M. (2002) A generalized uncertainty principle and sparse representation in pairs of $\mathbb{R}^N$ bases. *IEEE Trans. Inf. Theory*, **48**, 2558–2567.

[47] Gribonval, R. and Nielsen, M. (2003) Sparse representations in unions of bases. *IEEE Trans. Inf. Theory*, **49**, 3320–3325.

[48] Tropp, T. (2004) Topics in sparse approximation. PhD Thesis, University of Texas, Austin.

[49] Bickel, P. and Docksum, K. (2001) *Mathematical Statistics: Basic Ideas and Selected Topics* (2nd edn). Prentice-Hall, London.

[50] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1999) Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, **20**, 33–61.

[51] Sardy, S., Bruce, A. and Tseng, P. (2000) Block coordinate relaxation methods for nonparametric wavelet denoising. *J. Comput. Graph. Stat.*, **9**, 361–379.

[52] Antoniadis, A. and Fan, J. (2001) Regularization of wavelets approximations. *J. Am. Stat. Assoc.*, **96**, 939–963.

[53] Nikolova, M. (2000) Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.*, **61**, 633–658.

[54] Elad, M. (2005) Why simple shrinkage is still relevant for redundant representations? Technical report. Department of Computer Science, Technion.

[55] Elad, M., Matalon, B. and Zibulevsky, M. (2006) Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. Technical report. Department of Computer Science, Technion.

[56] Meng, X.L. and Rubin, D.B. (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80**, 267–278.

[57] Wu, C. (1983) On the convergence properties of the EM algorithm. *Ann. Stat.*, **11**, 95–103.

[58] Nikolova, M. (2005) Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares. *SIAM J. Multiscale Modeling Simul.*, **4**, 960–991.

[59] Combettes, P.L. and Wajs, V.R. (2005) Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Modeling Simul.*, **4**, 1168–1200.

[60] Meng, X.L. and Rubin, D.B. (1992) Recent extensions to the EM algorithm. In Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds.), *Bayesian Statistics*, pp. 307–320. Oxford University Press.

[61] Meng, X.L. (1994) On the rate of convergence of the ECM algorithm. *Ann. Stat.*, **22**, 326–339.

[62] Ciarlet, P.G. (1982) *Introduction à l'Analyse Numérique Matricielle et à l'Optimisation*. Masson, Paris.

[63] Glowinski, R., Lions, J. and Trémolières, R. (1976) *Analyse Numérique des Inéquations Variationnelles* (1st edn). Dunod, Paris.

[64] Tseng, P. (2001) Convergence of a block coordinate descent method for nondifferentiable minimizations. *J. Optim. Theory Appl.*, **109**, 457–494.

[65] Ueda, N. and Nakano, R. (1998) Deterministic annealing EM algorithm. *Neural Netw.*, **11**, 271–282.

[66] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. *Ann. Stat.*, **32**, 407–499.

[67] Donoho, D., Tsaig, Y., Drori, I. and Starck, J.-K. (2006) Sparse solution of underdetermined linear equations of stagewise orthogonal matching pursuit. *Technical Report. Dept. of Statistics*, Stanford University.

[68] Guleryuz, O. (2006) Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising. Part II: adpative algorithms. *IEEE Trans. Image Process*, **15**, 555–571.

[69] Buckheit, J. and Donoho, D. (1995) Wavelab and Reproducible Research. In Antoniadis, A. (ed.), *Wavelets and Statistics*. Springer.

[70] Guleryuz, O. (2006) Nonlinear approximation based image recovery using adaptive sparse reconstructions and iterated denoising. Part I: theory. *IEEE Trans. Image Process*, **15**, 539–554.

[71] Chan, R., Shen, L. and Shen, Z. (2005) A framelet-based approach for image inpainting. Technical report. The Chinese University of Hong Kong.

[72] Opial, Z. (1967) Weak convergence of the sequence of successive approximations for nonexpansive mappings. *Bull. Am. Math. Soc.*, **73**, 591–597.

[73] Daubechies, I., Defrise, M. and Mol, C.D. (2004) An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, **57**, 1413–1541.

8. APPENDIX

8.1. Proof of Proposition 1

Proof. Proof of (i) follows the same lines as in [52, 53]. When $\mathbf{\Phi}$ is a basis, the solution to Equation (6) is a componentwise minimization problem. When $z = 0$, it is clear that the unique minimizer is $\hat{\alpha} = 0$. Moreover, as the functional to be minimized is even-symmetric (owing to H), the solution function is odd, i.e. $\hat{\alpha}(-z) = -\hat{\alpha}(z)$. Thus, we study the problem only for $z > 0$. By differentiating (using H3), we have

$$\hat{\alpha} = z - \lambda\sigma^2\psi'(\hat{\alpha}). \tag{A.1}$$

Using a simple contradiction argument, α and z necessarily have the same sign. Indeed, assume that $\hat{\alpha} \leq 0$ when $z > 0$. But, by H1, the right-hand side of Equation (A.1) is positive, contradicting our initial assumption. Thus, if $z \leq \min_{\beta \geq 0} \beta + \lambda\sigma^2\psi'(\beta) = T$, the only positive solution to Equation (A.1) is $\hat{\alpha} = 0$. On the other hand, when $z > T$, Equation (A.1) may have many solutions. For ψ convex, it is trivial to see that $\alpha + \lambda\sigma^2\psi'(\alpha)$ is one-to-one for $z > T$, hence the solution is unique. Otherwise, owing to H4, the componentwise derivative written in Equation (A.1) has two possible zero

crossings on $(0, +\infty)$, but only the larger one is the minimum occurring at a strictly positive $\hat{\alpha}$. Finally, to ensure continuity of $\hat{\alpha}(z)$ at T, T is necessarily $\lambda\sigma^2\psi'_+(0)$.

Assertions (ii) and (iii) follow as a direct application of [57, Theorem 2], as since the surrogate functional $Q(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(t)})$ is continuous in both arguments (see H3). □

8.2. Proof of Proposition 2

Proof. Existence statement (i) is a consequence of coercivity, which is a sufficient condition for the inpainting problem to have a solution.

Uniqueness assertion (ii) follows as a consequence of [59, Proposition 5.15(ii)(a)], as $\mathcal{M}\boldsymbol{\Phi}$ is not necessarily bounded below.

To prove (iii), let us define the sequence of iterates

$$\mathbf{x}^{(t+1)} = \mathcal{D}_{\boldsymbol{\Phi},\lambda}\big(\mathbf{y}_{\text{obs}} + (\mathbf{I} - \mathcal{M})\mathbf{x}^{(t)}\big), \tag{A.2}$$

where $\mathcal{D}_{\boldsymbol{\Phi},\lambda} = \boldsymbol{\Phi} S_{\lambda\sigma^2} \boldsymbol{\Phi}^{\mathrm{T}}$ consists of applying the transform $\mathbf{T} = \boldsymbol{\Phi}^{\mathrm{T}}$, the componentwise shrinkage operator associated with Equation (14) with a parameter $\lambda\sigma^2$, and then reconstruct.

LEMMA 1. *If $\psi'(.)$ is non-decreasing, $\mathcal{D}_{\boldsymbol{\Phi},\lambda}$ is a non-expansive operator.*

Proof. Since the ℓ_2-norm is invariant under orthonormal transformations ($\boldsymbol{\Phi}$), it is sufficient to prove that, $\forall\, z, z' \in \mathbb{R}$

$$|\hat{\alpha}(z) - \hat{\alpha}'(z')| \leq |z - z'|. \tag{A.3}$$

This is just checked in all possible cases.

(i) If $|z| \leq T$ and $|z'| \leq T$, $|\hat{\alpha}(z) - \hat{\alpha}'(z')| = 0 \leq |\, z - z'|$.
(ii) Conversely, consider z and z' to have the same sign and $|z| > T$ and $|z'| > T$, the function $z(\alpha) = \alpha + \lambda\sigma^2\psi'(\alpha)$ is one-to-one monotonically increasing with derivative uniformly bounded below by 1, hence verifying Equation (A.3).
(iii) If $z > T$ and $z' < -T$, $|\hat{\alpha}(z) - \hat{\alpha}'(z')| = |z - z' - \lambda\sigma^2(\psi'(\alpha) - \psi'(\alpha'))| < |z - z'|$.
(iv) If $z > T$ and $|z'| < T$, by monotonicity of ψ':

$$\begin{aligned} |\hat{\alpha}(z)| - \hat{\alpha}'(z')| &= |z - \lambda\sigma^2\psi'(\alpha)| \\ &\leq |z - \lambda\sigma^2\psi'_+(0)| \\ &\leq |z - z'|. \end{aligned} \tag{A.4}$$

□

It is now straightforward to see that

$$\begin{aligned} \|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|^2 &\leq \big\|(\mathbf{I} - \mathcal{M})(\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)})\big\|^2 \\ &< \|\mathbf{x}^{(t)} - \mathbf{x}^{(t-1)}\|^2. \end{aligned} \tag{A.5}$$

The last strict inequality holds whenever some data is missing (i.e. avoiding the trivial identity mask). Furthermore, by standard properties of the EM (see Proposition 1(ii)), $Q(\mathbf{x}^{(t+1)}|\mathbf{x}^{(t)})$ and the penalized log-likelihood ℓ_ψ increase monotonically. Therefore, from the definition of the surrogate functional $Q(.|.)$ in Equation (9):

$$\begin{aligned} Q(\mathbf{x}^{(t+1)}|\mathbf{x}^{(t)}) &= \ell\ell(\mathbf{y}_{\text{o}}|\mathbf{x}^{(t+1)}) - \lambda\Psi\big(\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{x}^{(t+1)}\big) \\ &\quad - \frac{1}{2\sigma^2}\big\|(\mathbf{I} - \mathcal{M})\big(\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\big)\big\|_2^2 \\ &= \ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(t+1)}) \\ &\quad - \frac{1}{2\sigma^2}\big\|(\mathbf{I} - \mathcal{M})\big(\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\big)\big\|_2^2 \\ &\geq Q(\mathbf{x}^{(t)}|\mathbf{x}^{(t)}) = \ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(t)}). \end{aligned} \tag{A.6}$$

Thus,

$$\begin{aligned} &\sum_{t=0}^{N}\big\|(\mathbf{I} - \mathcal{M})\big(\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\big)\big\|_2^2 \\ &\quad \leq 2\sigma^2\sum_{t=0}^{N}\big(\ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(t+1)}) - \ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(t)})\big) \\ &\quad \leq 2\sigma^2\big(\ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(N)}) - \ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(0)})\big) \\ &\quad \leq -2\sigma^2\ell\ell_\psi(\mathbf{y}_{\text{o}}|\mathbf{x}^{(0)}) < +\infty. \end{aligned} \tag{A.7}$$

It follows that the series $\sum_{t=0}^{N}\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2^2$ is bounded uniformly in N, so that $\sum_{t=0}^{\infty}\|\mathbf{x}^{(t+1)} - \mathbf{x}^{(t)}\|_2^2$ converges. Thus, the sequence of iterates is asymptotically regular. Finally, as the set of fixed points is non-empty (coercivity), $\mathcal{D}_{\boldsymbol{\Phi},\lambda}$ is non-expansive and the sequence $\mathbf{x}^{(t)}$ is asymptotically regular. By application of Opial's theorem [72], statement (iii) follows. The proof provided here bears similarities with the one given in [73].

To prove statement (iv), it is sufficient to identify our problem with the one considered by Combettes and Wajs [59, Problem 5.13], and their fixed-point iteration [59, Problems (5.24) and (5.25)] with ours. It turns out that the shrinkage operator $S_{\lambda\sigma^2}$ in Equation (14) is the Moreau proximity operator associated with the convex function ψ, the lossy operator $\mathcal{M}$ is a bounded linear operator, and $\boldsymbol{\Phi}$ is a bijective bounded linear operator. Thus, strong convergence follows from [59, Corollary 5.19]. □

8.3. Proof of Proposition 3

Proof. The proof for frames is less straightforward that for a basis. To prove our result, we invoke a theorem due to [59, Theorem 5.5]. For the sake of completeness and the reader's convenience, we here state in full some of their results relevant to our setting, but without a proof.

THEOREM 1 (Combettes and Wajs [59]). *Consider the optimization problem*

$$\underset{\alpha}{\text{minimize}}\, \frac{1}{2}\|\mathbf{y}-\mathbf{A}\boldsymbol{\alpha}\|_2^2 + f(\mathbf{W}^*\boldsymbol{\alpha}), \tag{A.8}$$

where $\mathbf{A}$ *is a bounded linear operator mapping a Hilbert space to another*, $\mathbf{W}$ *is a bijective bounded linear operator* (*its adjoint* $\mathbf{W}^*$ *is equal to its inverse*) *and* $f(.)$ *is a lower semi-continuous proper convex function. Let* $G \neq \emptyset$ *be the set of solutions to Problem* 26. *Let* $\{\mu_t,\ t \in \mathbb{N}\}$ *be a sequence in* $(0, +\infty)$ *such that* $0 < \inf_t\ \mu_t \leq \sup_t\ \mu_t < 2/\|\mathbf{A}\|_2^2$, *let* $\{\beta_t,\ t \in \mathbb{N}\}$ *be a sequence in* $(0, 1]$ *such that* $\inf_t\ \beta_t > 0$, *and let* $\{a_t,\ t \in \mathbb{N}\}$ *and* $\{b_t,\ n \in \mathbb{N}\}$ *be sequences in a Hilbert space such that* $\sum_t \|a_t\| < +\infty$ *and* $\sum_t \|b_t\| < +\infty$. *Fix* $\boldsymbol{\alpha}_0$, *and define the sequence of iterates*.

$$\begin{aligned}\boldsymbol{\alpha}^{(t+1)} = \boldsymbol{\alpha}^{(t)} + \beta_t\Big(\Big(\mathbf{W}\circ \operatorname{prox}_{\mu_t f}\circ\mathbf{W}^*\Big) \\ \times\big(\boldsymbol{\alpha}^{(t)} + \mu_t\big(\mathbf{A}^*\big(\mathbf{y}-\mathbf{A}\boldsymbol{\alpha}^{(t)}\big)\big) - b_t\big) + a_t - \boldsymbol{\alpha}^{(t)}\Big),\end{aligned} \tag{A.9}$$

where $\operatorname{prox}_{\mu t, f}$, *is the proximity operator of* f.[4]

(i) *Problem* 26 *has at least one solution if* f *is coercive*.
(ii) *Problem* 26 *has at most one solution if* f *is strictly convex or* $\mathbf{A}$ *is injective*.
(iii) $\{\mathbf{x}^{(t)},\ t \geq 0\}$ *converges weakly to a point* $\hat{\boldsymbol{\alpha}} \in G$.
(iv) $\{\mathbf{x}^{(t)},\ t \geq 0\}$ *converges strongly to* $\hat{\boldsymbol{\alpha}} \in G$ *in each of the following cases*.
  (a) *The interior of* G *is non-empty*.
  (b) f *satisfies Condition* 3.2 *in* [59].

We now reformulate our inpainting problem as a linear inverse problem regularized over frames. The problem under consideration can be stated as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \frac{1}{2\sigma^2}\|\mathbf{y}_{\text{obs}} - \mathcal{M}\boldsymbol{\Phi}\boldsymbol{\alpha}\|_2^2 + \lambda\sum_l \psi(\boldsymbol{\alpha}_l). \tag{A.10}$$

Thus, our inpainting inverse problem is a special case of Problem (26) with $\mathbf{A} = \mathcal{M}\boldsymbol{\Phi}$, which is a linear bounded measurement operator, but not injective, $\mathbf{W} = \mathbf{I}$, and $f(\boldsymbol{\alpha}) = \sum_l \psi(\boldsymbol{\alpha}_l)$. The first two assertions (existence and uniqueness) of Proposition 3 then follow from Theorem 1(i) and (ii).

To determine the proximity operator $\operatorname{prox}_{\mu_t, f}$, we first notice that in our case $\mathbf{W}: \boldsymbol{\alpha} \mapsto (\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{e}_l)_{l=1,\ldots,L}$, $(\mathbf{e}_l)_{l=1,\ldots,L}$ is the canonical basis of $\mathbb{R}^L$. Hence, following the same proof as in Proposition 1, the proximity operator $\operatorname{prox}_{\mu_t, f}$ amounts to the shrinkage operator $S_{\mu_t\lambda\sigma^2}$ given in Equation (14), applied componentwise to each $\boldsymbol{\alpha}_l$.

[4] See [59, Section 2.3] for more details about proximal calculus.

Furthermore, by definition of frames, there exist two constants A, $B > 0$ such that $\forall \mathbf{x} \in \mathcal{H}$

$$A \leq \frac{\mathbf{x}^{\mathrm{T}}(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}})\mathbf{x}}{\|\mathbf{x}\|_2^2} \leq B. \tag{A.11}$$

That is, any Rayleigh quotient of the frame operator $\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}$ must lie between A and B. It follows from the Courant–Fischer theorem that each eigenvalue of the frame operator lies in the interval $[A, B]$, and its spectral radius $\rho(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}) = B$. Thus,

$$\begin{aligned}\|\mathbf{A}\|_2^2 = \|\mathcal{M}\boldsymbol{\Phi}\|_2^2 = \|\mathcal{M}\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\mathcal{M}\|_2 \\ \leq \|\mathcal{M}\|_2^2\|\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\|_2 = \rho(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}) = B,\end{aligned} \tag{A.12}$$

where the inequality becomes an equality for tight frames. Hence, taking $0 < \mu_t = \mu < 2/B$, $\beta_t \equiv 1$, $a_t \equiv 0$ and $b_t \equiv 0$ in Equation (A.9) our iteration in Equation (15) follows. Statement (iii) of Proposition 3 is then a straightforward application of Theorem 1(iii).

For the inpainting algorithm to converge in a strong sense with frames, it is sufficient that ψ be strongly convex (see [59, Proposition 3.6(vii)]. However, this kind of penalty is not relevant for producing sparse solutions. If $\psi(\alpha) = |\alpha|^p$, $p \in [1,2]$, we conclude with the same argument as in Proposition 2(iv) for a basis, with $(\mathbf{e}_l)_{l=1,\ldots,L}$ the canonical basis of $\mathbb{R}^L$. □

8.4. Proof of Proposition 4

Proof. Note first that the solution existence and uniqueness assertions can be stated as before using coercivity and strict convexity arguments.

The convergence proof follows, after identifying our problem with the one considered by the author in [64]. Then using the same notation as in [64]

$$\begin{aligned}f_0(\boldsymbol{\alpha}) &= \frac{1}{2\sigma^2}\left\|\mathbf{y}_{\text{obs}} - \mathcal{M}\sum_{k=1}^{K}\boldsymbol{\Phi}_k\boldsymbol{\alpha}_k\right\|_2^2, \\ f_k(\boldsymbol{\alpha}_k) &= \lambda_k\Psi(\boldsymbol{\alpha}_k), \qquad \forall k = 1,\ldots,K.\end{aligned}$$

It is not difficult to see that f_0 has a non-empty open domain and is continuously differentiable on its domain. Thus, f_0 satisfies Assumption A.1 in [64]. Moreover, by assumption on ψ, the objective functional $f_0(\boldsymbol{\alpha}) + \sum_k f_k(\boldsymbol{\alpha}_k)$ is continuous on its effective domain, with bounded level sets, and is convex (hence pseudoconvex). Thus, by Tseng [64 Lemma 3.1 and Theorem 4.1], the convergence statement follows. The accumulation point of the iteration is the unique global minimizer if ψ is strictly convex. □