SPARSE REPRESENTATIONS AND BAYESIAN IMAGE INPAINTING

M.J. Fadili

GREYC UMR CNRS 6072
10 Bd Maréchal Juin 14050 Caen France

J-L. Starck

DAPNIA/SEDI-SAP, Service d'Astrophysique
91191 Gif-sur-Yvette France

ABSTRACT

Representing the image to be inpainted in an appropriate sparse dictionary, and combining elements from bayesian statistics, we introduce an expectation-maximization (EM) algorithm for image inpainting. From a statistical point of view, the inpainting can be viewed as an estimation problem with missing data. Towards 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 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 for a wide class of penalties. Particularly, we establish that it converges in a strong sense, and give sufficient conditions for convergence to a local or a global minimum. Compared to its competitors, this algorithms allows a high degree of flexibility to recover different structural components in the image (piece-wise smooth, curvilinear, texture, etc). We also describe some ideas to automatically find the regularization parameter.

1. INTRODUCTION

Inpainting is to restore missing image information based upon the still available (observed) cues. 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 desocclusion problem. The classical image inpainting problem can be stated as follows. Suppose the ideal complete image X defined on a finite domain Ω (the plane), and its degraded version (but not completely observed) Y. The observed (incomplete) image $Y_{\rm obs}$ is the result of applying the lossy operator $\mathcal{M}$ on Y:

$$\mathcal{M} : Y \mapsto Y_{\rm obs} = \mathcal{M}\left[Y\right] = \mathcal{M}\left[X \odot \varepsilon\right] \qquad (1)$$

where $\odot$ is any composition of two arguments (e.g. '+' for additive noise, etc), ε is the noise. $\mathcal{M}$ is defined on $\Omega \setminus E$, 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). Inpainting is to recover X from $Y_{\rm obs}$ which is an inverse ill-posed problem.

Recent wave of interest in inpainting was started from the pioneering work of [1], where applications in the movie industry, video, and art restoration were unified. These authors proposed nonlinear PDE model for inpainting. Following their work, [2] then systematically investigated inpainting based on the Bayesian and (possibly hybrid) variational principles with different penalizations (TV, l_1 norm on wavelets coefficients). Many other authors have also proposed inpainting algorithms under the variational/PDE framework. More recently, [3] introduced a novel inpainting algorithm that is capable of reconstructing both texture and cartoon image contents, i.e. $X = \Phi\boldsymbol{\alpha}$, where Φ is a dictionary of sparse transforms (e.g. curvelets for cartoon and local cosines for locally stationary textures). This algorithm is a direct extension of the MCA (Morphological Component Analysis), designed for the separation of an image into different semantic components [4, 5].

Combining elements from statistics and harmonic analysis theories, we here introduce an EM algorithm for image inpainting based on a penalized maximum likelihood formulated using linear sparse representations, i.e. $X = \Phi\boldsymbol{\alpha}$, where the image X is supposed to be efficiently by the 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 previous iteration, and then 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, 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 to investigate the convergence behavior of the inpainting algorithm. Some results are finally shown to illustrate our algorithm.

2. PENALIZED MLE WITH MISSING DATA

2.1. Problem formulation

Suppose that the an image has n pixels. First, let's ignore the missing data mechanism and write the complete

n-dimensional observation vector (by simple reordering) Y as:

$$Y = \Phi\boldsymbol{\alpha} + \varepsilon, \varepsilon \sim \mathcal{N}(0, \sigma^2) \tag{2}$$

Φ is a $n \times p$ matrix corresponding to a sparse representation (possibly overcomplete $p \geq n$). Estimating X from Y can be accomplished using the penalized maximum likelihood estimator (PMLE):

$$\hat{X} = \arg\min_X -\ell\ell\left(Y|X\right) + \log p_X(x) \tag{3}$$

As X is supposed to be sparsely decomposed in the chosen dictionary. The MAP/PMLE estimation problem can then be expressed in terms of the decomposition coefficients $\boldsymbol{\alpha}$, which gives, for additive white Gaussian noise with known variance σ^2:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{2\sigma^2}\|Y - \Phi\boldsymbol{\alpha}\|_2^2 + \lambda\Psi\left(\boldsymbol{\alpha}\right) \tag{4}$$

where $\Psi(\boldsymbol{\alpha})$ is a penalty function promoting reconstruction with low complexity taking advantage of sparsity. In the sequel, we additionally assume that the prior associated to $\Psi(\boldsymbol{\alpha})$ is separable (i.e. coefficients independence). Hence,

$$\Psi(\boldsymbol{\alpha}) = \sum_{l=1}^{L} \psi(|\alpha_l|) \tag{5}$$

2.2. Redundant sparse representations

Suppose $X \in \mathcal{H}$ a Hilbert space. An $\sqrt{n} \times \sqrt{n}$ image X can be written as the superposition of elementary functions $\phi_\gamma(u, v)$ (atoms) parameterized by γ s.t. (Γ is denumerable):

$$X(u, v) = \sum_{\gamma \in \Gamma} \boldsymbol{\alpha}_\gamma \phi_\gamma(u, v),\ \phi_\gamma \in \mathcal{L} \tag{6}$$

where the atoms $\{\phi_l\}_{l=1,\ldots,L}$ are normalized to a unit norm. The forward transform is defined by $\Phi = [\phi_1 \ldots \phi_L] \in \mathbb{R}^{n \times L}$, $\text{Card}\ \Gamma = L \gg N$ (*union* of incoherent bases, of frames or tight frames), and Φ has a Moore-Penrose generalized-inverse (Φ^+). 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).

2.3. The EM algorithm

Let's now turn to the missing data case and let's write $Y = (Y_{\text{obs}}, Y_{\text{miss}})$, with $Y_{\text{miss}} = \{y_i\}_{i \in I_m}$ is the missing data, and $Y_{\text{obs}} = \{y_i\}_{i \in I_o}$. The incomplete observations do not contain all information to apply standard methods to solve (4) and get the PMLE of $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \sigma^2)^T \in \Theta \subset \mathbb{R}^p \times \mathbb{R}^{+*}$. Nevertheless, the EM algorithm can be applied to iteratively reconstruct the missing data and then solve (4) for the new estimate. The estimates are iteratively refined until convergence.

The E step
This steps computes the conditional expectation of the penalized log-likelihood of complete data, given Y_{obs} and current parameters $\boldsymbol{\theta}^{(t)} = \left(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\theta}'^{(t)}\right)^T$:

$$\begin{aligned} Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) &= \mathsf{E}\left[\ell\ell\left(Y|\boldsymbol{\theta}\right) - \lambda\Psi\left(\boldsymbol{\alpha}\right)|Y_{\text{obs}}, \boldsymbol{\theta}^{(t)}\right] \\ &= \mathsf{E}\left[\underbrace{\ell\ell\left(Y|\boldsymbol{\theta}\right)}_{\sim \text{ Data fidelity}}|Y_{\text{obs}}, \boldsymbol{\theta}^{(t)}\right] - \lambda \underbrace{\Psi\left(\boldsymbol{\alpha}\right)}_{\text{Prior penalty}} \end{aligned} \tag{7}$$

For regular exponential families, the E steps reduces to finding the expected values of the sufficient statistics of the complete data Y given observed data Y_{obs} 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:

$$\begin{aligned} y_i^{(t)} &= \mathsf{E}\left(y_i|\Phi, Y_{\text{obs}}, \boldsymbol{\alpha}^{(t)}, \sigma^{2^{(t)}}\right) = \begin{cases} y_{obs_i} & \text{for observed data, } i \in I_o \\ \left(\Phi\boldsymbol{\alpha}^{(t)}\right)_i & \text{for missing data, } i \in I_m \end{cases} \\ \text{and } \mathsf{E}\left(y_i^2|\Phi, Y_{\text{obs}}, \boldsymbol{\alpha}^{(t)}, \sigma^{2^{(t)}}\right) &= \begin{cases} y_{obs_i}^2 & i \in I_o \\ \left(\Phi\boldsymbol{\alpha}^{(t)}\right)_i^2 + \sigma^{2^{(t)}} & \imath \in I_m \end{cases} \end{aligned} \tag{8}$$

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

$$\boldsymbol{\theta}^{(t+1)} = \arg\min_{\boldsymbol{\theta} \in \Theta}\ -Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) \tag{9}$$

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

$$\sigma^{2^{(t+1)}} = \frac{1}{n}\left[\sum_{i \in I_o}\left(y_i - x_i^{(t)}\right)^2 + (n - n_o)\sigma^{2^{(t)}}\right] \tag{10}$$

where $n_o = \text{tr}\ \mathcal{M} = \text{Card}\ I_o$ is the number of observed pixels. Note that at convergence, we have $\hat{\sigma}^2$ is the noise variance estimate inside the mask (i.e. with observed pixels). The update equation of X^{t+1} is more complicated and will be detailed hereafter.

2.4. The ECM inpainting algorithm

The Cyclic EM (ECM) M-step is accomplished by sequentially cycling between the atoms in the dictionary and minimizing with respect to each α_l keeping the other coefficients fixed.

Require: Observed image Y_{obs} and a mask $\mathcal{M}$, convergence threshold δ,

1: **repeat**
2: ***E Step***
3: Update the image estimate:

$$Y^{(t)} = Y_{\text{obs}} + (I - \mathcal{M})X^{(t)} \tag{11}$$

CM Step
for Each column l of Φ **do**
  Compute the transform coefficient $\phi_l^T \left(Y^{(t)} - \Phi\boldsymbol{\alpha}^{(t)}\right) + \alpha_l^{(t)}$,
  Apply the shrinkage operator $\mathcal{D}_\lambda$ associated to $\psi(.)$ (e.g. soft thresholding for $\psi(|\alpha|) = |\alpha|$) to this coefficient to obtain $\alpha_l^{(t+1)}$,
end for
Update $X^{(t+1)} = \Phi\boldsymbol{\alpha}^{(t+1)}$,
Update ${\sigma^2}^{(t+1)}$ according to (10).
until Convergence, i.e. $\left\|X^{(t+1)} - X^{(t)}\right\|_2 \leq \delta$

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

2.5. Convergence results of the ECM inpainting

We now summarize the main features of the above algorithm in the following theorem.

Theorem 1 *Suppose that:*

H 1. *ψ is even-symmetric, nonnegative and nondecreasing on $[0, +\infty)$, and $\psi(0) = 0$.*

H 2. *ψ is twice differentiable on $\mathbb{R} \setminus \{0\}$ but not necessarily convex.*

H 3. *ψ is continuous on $\mathbb{R}$, it is not necessarily smooth at zero and admits a positive right derivative at zero $p'_+(0) = \lim_{h \to 0^+} \frac{p(h)}{h} > 0$ which can be finite or not.*

H 4. *The function $\alpha + \lambda\psi'(\alpha)$ is unimodal on $(0, +\infty)$.*

H 5. *The columns of Φ are normalized to a unit ℓ_2 norm.*

Then,

(i) The sequence of observed penalized likelihood converges monotonically to some $\ell\ell^*$.

(ii) All limit points of the ECM inpainting sequence $\{X^{(t)}, t \geq 0\}$ are stationary points of the penalized likelihood.

(iii) The sequence of iterates is asymptotically regular, i.e. $\left\|X^{(t+1)} - X^{(t)}\right\| \to 0$.

(iv) If ψ is unimodal, then any inpainting sequence converges to the unique minimizer.

Sketch of Proof: Statements (i)-(ii) follow from continuity of $\psi(.)$ and classical results on the ECM [6, 7]. Statement (iii) is a consequence of assumptions H1-H4, yielding that point-wise minimization with respect to each α_l is single-valued. Statement (iv) follows from convexity of $\psi(.)$ [6, 7]. ■

2.6. Computational complexity

The computational complexity of the above algorithm is dominated by the multiplication by the columns of Φ used in the M step, which is typically $O(LN)$ (particularly for redundant dictionaries $L \gg N$). Thus, for most transforms popular in harmonic analysis used with large scale image processing applications, this would be of prohibitive computational burden. Therefore, the question is do fast solution exist for such a problem that can reduce the complexity reasonably for usual transforms ? Fortunately, the answer is yes provided that the dictionary is a union of bases ($\Phi_k\Phi_k^T = \mathbf{I}$) or a union of tight frames ($\Phi_k\Phi_k^T = c\mathbf{I}$). The CM steps of the above algorithm can then be rewritten as a parallel updating scheme:

for Each transform $k \in [1, K]$ in the dictionary **do**
  Compute the coefficients $\Phi_k^T\left(Y_{\text{obs}} - \mathcal{M}X^{(t)}\right) + \boldsymbol{\alpha}_k^{(t)}$,
  Apply the shrinkage operator $\mathcal{D}_\lambda$ to these coefficients to obtain $\boldsymbol{\alpha}_k^{(t+1)}$,
end for

where applying Φ_k and Φ_k^T corresponds to the inverse and forward transforms (up to a scalar for tight frames). Consequently, the complexity becomes $O(Kg(N))$, $K \ll L$, where $g(N)$ is typically $O(N)$ or $O(N\log(N))$ for most usual transforms. Furthermore, the conclusions of Theorem 1 are still valid.

2.7. Choice of the regularization parameter

So far, we have characterized the solution $\hat{X}$ for a particular choice of λ. This choice is a challenging task. One attractive solution is based upon the following observation. At early stages of the algorithm, posterior distribution of α is unreliable because of missing data. One should then consider a large value of λ (∞ or equivalently $\|\Phi^+ Y\|_\infty$ to favor the penalty term. λ is then incrementally decreased (according to some schedule) to find and trace optimal solutions $\hat{X}(\lambda)$ for all $\lambda \geq k\sigma$ (to reject the noise). This procedure has a flavor of deterministic annealing, where the regularization parameter parallels the temperature. It can be also seen as the basis of a homotopy continuation method.

3. EXPERIMENTAL RESULTS

The ECM inpainting algorithm was applied to several synthetic and real degraded images, from which we present few examples.Fig.1 depicts an example on Lena where 80% pixels were missing. The dictionary contained the curvelet transform and the convex l_1 penalty was used. The threshold parameter was fixed to the universal 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 are almost impossible to distinguish in the masked image.

Fig. 1. Example with Lena. Dictionary: curvelets, penalty: l_1, input $SNR = 25dB$, 80% pixels missing.

Fig. 2. Examples with Lena and Barbara. Dictionary: curvelets+LDCT, penalty: l_1, input $SNR = 30dB$, 20% pixels missing.

To further illustrate the power of the ECM inpainting algorithm, we applied it to the Barbara textured image. As stationary textures are efficiently represented by the local DCT, the dictionary contained both the curvelet (for the geometry part) and the LDCT transforms. Again, the l_1 penalty was used. The result is portrayed in Fig.2. The algorithm is not only able to recover the geometric part (cartoon), but particularly performs well inside the difficult textured areas, e.g. trousers.

The algorithm was finally applied to a real photograph of a parrot, where we wanted to remove the grid of the cage to virtually free the bird. The mask of the grid was manually plotted. The result is shown in Fig.3. Here, the dictionary contained the undecimated DWT and LDCT, penalty: l_1. For comparative purposes, inpainting results of a PDE-based method [8] are reported. The main differences between the two approaches are essentially concentrated the "textured" area in the vicinity of the parrot's eye.

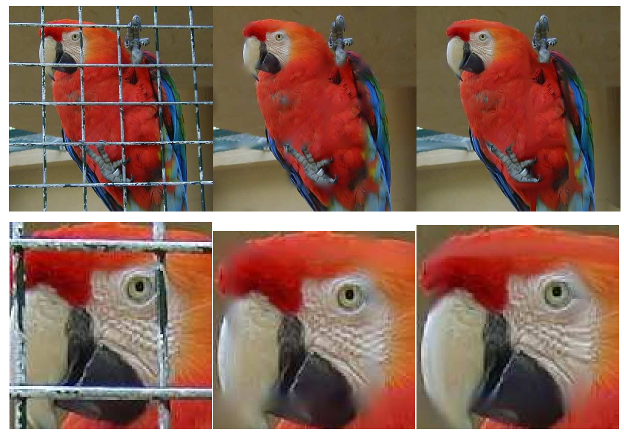

Fig. 3. Original (left), ECM inpainting (middle) and PDE inpainting (right).

4. CONCLUSION

A novel, fast and flexible inpainting algorithm has been presented. Its theoretical properties were also derived. Several interesting perspectives of this work are under investigation. We can cite the extension to any dictionary and formal investigation of the influence of the regularization parameter on the convergence the algorithm (path following/homotopy continuation). Extension to multi-valued images (e.g. hyper-spectral data) is also an important aspect that is the focus of our current research.

5. REFERENCES

[1] M. Bertalmìo, G. Sapiro, V. Caselles, and C. Ballester., "Image inpainting," in *SIGGRAPH 2000*, New Orleans, USA, 2000, vol. 34, pp. 417–424.

[2] T. Chan and J. Shen, "Mathematical models for local non-texture inpainting," *SIAM J. Appl. Math*, vol. 62, no. 3, pp. 1019–1043, 2001.

[3] M. Elad, J.-L. Starck, P. Querre, and D. Donoho, "Simultaneous cartoon and texture image inpainting," *submitted to ACHA*, 2004.

[4] J.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representatntions and variational approach," *IEEE Trans. Im. Proc.*, p. in press, 2004.

[5] J.-L. Starck, M. Elad, and D.L. Donoho, "Redundant multiscale transforms and their application for morphological component analysis," *Advances in Imaging and Electron Physics*, vol. 132, 2004.

[6] X. L. Meng and D. B. Rubin, "Maximum likelihood estimation via the ecm algorithm: A general framework," *Biometrika*, vol. 80, no. 2, pp. 267–278, 1993.

[7] C. Wu, "On the convergence properties of the em algorithm," *Ann. of Stat.*, vol. 11, pp. 95–103, 1983.

[8] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with pde's : A common framework for different applications," *IEEE Trans. PAMI*, vol. 27, no. 27, pp. 506–517, 2005.