EM Algorithm for Sparse Representation-based Image Inpainting

Jalal M. Fadili, Jean-Luc Starck

EM ALGORITHM FOR SPARSE REPRESENTATION-BASED IMAGE INPAINTING

M.J. Fadili

GREYC UMR CNRS 6072
10 Bd Maréchal Juin 14050 Caen France

J-L. Starck[*]

CEA/Saclay Service d'Astrophysique
91191 Gif-sur-Yvette France

ABSTRACT

We introduce an expectation-maximization (EM) algorithm for image inpainting based on a penalized likelihood formulated using linear sparse representations. Taking advantage of the sparsity of representations, a regularization through a prior penalty is imposed on the reconstructed coefficients. From a statistical point of view, the inpainting can be viewed as an estimation problem with missing data. The EM framework is a general iterative algorithm for ML estimation in such situations. 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.

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. $X = \Phi\alpha$, where Φ is a dictionary of sparse transform (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].

Combining elements from statistics and harmonic analysis theories, we here introduce an EM algorithm [5] for image inpainting based on a penalized maximum likelihood formulated using linear sparse representations, i.e. $X = \Phi\alpha$, where the image X is supposed to be efficiently by the atoms in the dictionary. Taking advantage of the sparsity of representations, a regularization through a 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. The EM framework is a very general iterative algorithm for ML estimation 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 [5, 6]. 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. It also turns out that there are several connections between our EM algorithm and the BCR-based inpainting algorithm as proposed recently by [3]. 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 [7], 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-

[*]The author performed the work while at the Department of Statistics Stanford University, USA.

dimensional observation vector (by simple reordering) Y as:

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

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) $\hat{X} = \arg\min_X -\ell\ell(Y|X) + \log p_X(x)$. This estimator can also be viewed as a MAP Bayesian estimator where $p_X(x)$ is the prior density function. 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 α, which gives, for additive white gaussian noise with known variance σ^2:

$$\hat{\alpha} = \arg\min_{\alpha} \frac{1}{2\sigma^2}\|Y - \Phi\alpha\|_2^2 + \lambda\Psi(\alpha) \quad (3)$$

where $\Psi(\alpha)$ is a penalty function promoting reconstruction with low complexity taking advantage of sparsity. Standard assumptions about the penalty (or -log-prior) is to consider that it is continuous, even-symmetric and strictly increasing on $(0, +\infty)$ (but not necessarily convex). A bunch of work departing from this equation has been proposed in the literature for the complete data setting (the so-called basis-pursuit problem).

2.2. 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 (3) and get the PMLE of $\boldsymbol{\theta} = (\alpha^T, \sigma^2)^T \in \Theta \subset \mathbb{R}^p \times \mathbb{R}^{+*}$. Nevertheless, the EM algorithm can be applied to iteratively reconstruct the missing data and then solve (3) for the new estimate. The estimates are iteratively refined until convergence. Recall that the EM algorithm is a means of obtaining MAP/PMLE estimates (of which maximum likelihood is a particular case) of a parameter in cases where the PMLE is hard to obtain. The (Bayesian) EM algorithm will then produce a sequence of estimates alternating between two steps:

- **E-step** Computes the conditional expectation of the log-likelihood of the complete data, given the observed data and the current estimate $\boldsymbol{\theta}^{(t)}$, by defining the surrogate function:

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \mathsf{E}\left[\ell\ell\left(Y|\boldsymbol{\theta}\right)|Y_{\text{obs}}, \boldsymbol{\theta}^{(t)}\right] - \lambda\Psi(\alpha) \quad (4)$$

- **M-step** Updates the estimates according to:

$$\boldsymbol{\theta}^{(t+1)} = \arg\min_{\boldsymbol{\theta} \in \Theta} -Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) \quad (5)$$

2.2.1. The E step

For regular exponential families, it is known that the E steps involves finding the expected values of the sufficient statistics of the complete data Y given observed data Y_{obs} and the estimate of $\alpha^{(t)}$ and $\sigma^{2^{(t)}}$ [5, 6]. Then, as the noise is zero-mean white gaussian, the E-step (4) reduces to calculating the conditional expected values and the conditional expected squared values of the missing data, that is:

$$y_i^{(t)} = \mathsf{E}\left(y_i|\Phi, Y_{\text{obs}}, \alpha^{(t)}, \sigma^{2^{(t)}}\right) = \begin{cases} y_{obs_i} & \text{for observed data, } i \in I_o \\ \Phi_i^T\alpha^{(t)} & \text{for missing data, } i \in I_m \end{cases}$$

$$\text{and } \mathsf{E}\left(y_i^2|\Phi, Y_{\text{obs}}, \alpha^{(t)}, \sigma^{2^{(t)}}\right) = \begin{cases} y_{obs_i}^2 & i \in I_o \\ \left(\Phi_i^T\alpha^{(t)}\right)^2 + \sigma^{2^{(t)}} & \imath \in I_m \end{cases} \quad (6)$$

2.2.2. The M step

This step consists in maximizing the penalized surrogate function of (4) with the missing observations replaced by their estimates in the E step at iteration t, that is:

$$\boldsymbol{\theta}^{(t+1)} = \arg\min_{\alpha, \sigma^2} \frac{1}{2\sigma^2}\|Y^{(t)} - \Phi\alpha\|_2^2 + \lambda\Psi(\alpha) + \frac{n}{2}\log 2\pi\sigma^2 \quad (7)$$

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

$$X^{(t+1)} = \Phi\mathcal{D}\Phi^+Y^{(t)} \quad (8)$$

$$\sigma^{2^{(t+1)}} = \frac{1}{n}\left[\sum_{i \in I_o}\left(y_i - x_i^{(t)}\right)^2 + (n - n_o)\sigma^{2^{(t)}}\right] \quad (9)$$

where $n_o = \text{tr}\,\mathcal{M} = \text{Card}\,I_o$ is the number of observed pixels. $\mathcal{D}$ denotes whichever estimation operation, associated to the penalty function $\Psi(\alpha)$, applied to the expansion coefficients in Φ. Φ^+ is the Moore-Penrose generalized inverse. Note that at convergence, we have:

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

which is the noise variance inside the mask (i.e. with observed pixels). If the noise variance is known in advance, the re-estimation of σ^2 in the M step can be ignored.

A very important feature of the M step is that it involves a denoising operation (8) depending on the choice of the penalty function. For example, under the l_1 norm penalization and an orthogonal dictionary, $\alpha^{(t+1)}$ can be simply estimated using the well known soft thresholding scheme. This can also be extended to redundant representations as we will see later. Other prior penalties will lead to different estimation rules in the M-Step.

2.3. Extension to highly redundant sparse representations

Our goal is to extend the EM algorithm above while preserving its convergence properties, e.g. monotonic decrease of the -log-posterior function (3). The algorithm is dedicated to dictionaries whose individual transformations are either bases, frames or tight frames (e.g. a dictionary composed of the curvelet, wavelet, local DCT transforms, etc).

Suppose that $\Phi = (\Phi_1, \ldots, \Phi_K)$, where Φ_k is the operator associated with the transformation k. Hence, the Generalized EM [5, 6] M-step is accomplished by cycling (either randomly or sequentially) between the individual transforms and minimizing with respect to each α_k keeping the other coefficients fixed, that is:

$$\alpha_k^{(t+1)} = \arg\min_{\alpha_k} \frac{1}{2\sigma^2} \left\| \left(Y^{(t)} - \sum_{k' \neq k} \Phi_{k'} \alpha_{k'}^{(t)} \right) - \Phi_k \alpha_k \right\|_2^2 + \lambda \Psi(\alpha_k) \tag{11}$$

where we additionally assumed that $\Psi(\alpha)$ is separable (e.g. l_1 norm), and σ^2 is fixed for simplicity. Note the similarities with the BCR algorithm. Moreover, compared to the EM algorithm, it is easy to verify that this GEM algorithm satisfies the weaker condition:

$$Q\left(\alpha^{(t+1)}|\alpha^{(t)}\right) \geq Q\left(\alpha^{(t)}|\alpha^{(t)}\right) \tag{12}$$

Owing to the general convergence properties of the EM (and the GEM), the above condition will also entail an increase of the penalized likelihood.

2.4. The GEM inpainting algorithm

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

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

$$Y^{(t)} = Y_{\text{obs}} + (I - \mathcal{M}) X^{(t)} \tag{13}$$

4: ***M Step***
5: **for** Each transformation k in the dictionary **do**
6: Calculate the transform coefficients $\Phi_k^+ \left(Y^{(t)} - \sum_{k' \neq k} \Phi_{k'} \alpha_{k'}^{(t)} \right)$,
7: Apply the estimation operator $\mathcal{D}$ (e.g. soft thresholding) to these coefficients and get $\alpha_k^{(t+1)}$,
8: **end for**
9: Update $X^{(t+1)} = \sum_k \Phi_k \alpha_k^{(t+1)}$,
10: Update $\sigma^{2^{(t+1)}}$ according to (9).
11: **until** Convergence, i.e. $\frac{\left\| X^{(t+1)} - X^{(t)} \right\|_2}{\left\| X^{(t)} \right\|_2} \leq \delta$

The computational complexity of this algorithm is dominated by that of the transformations used in the M step. More interestingly, if the update equation of the E step is replaced by:

$$r^{(t)} = Y_{\text{obs}} - \mathcal{M} X^{(t)} \tag{14}$$

where $r^{(t)}$ are the residuals inside the mask, then step 6 of the above algorithm will simply reduce to calculating the coefficients of $\Phi_k^+ \left(r^{(t)} + \Phi_k \alpha_k^{(t)} \right)$, which involves calculating only the transformation corresponding to Φ_k for each k. This considerably reduces the computational load of the M step.

A formidable impact of this remark is that it turns out that these modified update equations are exactly the same as those of the MCA-based inpainting recently proposed by [3] for simultaneous cartoon and texture recovery. Therefore, the EM framework gives a principled way to establish formally the intuitive idea that missing samples can be recovered based on sparse representations. It also generalizes it to any penalty function. However, an important difference between the GEM and the MCA-based algorithms is the evolution of the regularizing parameter λ. Here, we considered it as a fixed hyperparameter directly embedded in the penalty function, while the MCA decreases it according to a given schedule (linear or exponential). This would be expected to confer to the MCA a greater chance to converge to one of the global minima, but no proof of this has been established yet. On the other hand, the GEM can handle the case where σ^2 must be adaptively estimated, and owing to its well known theoretical properties, the GEM algorithm allows to investigate the convergence behavior of the inpainting algorithm.

3. CONVERGENCE ANALYSIS OF THE EM INPAINTING

We here characterize the existence and unicity of the solutions of the EM-based image inpainting. The convergence study carried out in this section is essentially based on the seminal results of [5] and revisited by [7].

3.1. Convergence to a stationary point

The EM provides a sequence of estimates whose penalized -log-likelihood is non-decreasing and bounded. Then, following [7, Theorem 2], a sufficient condition which guarantees the convergence of the sequence of EM inpainting to a stationary point is that $Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right)$ is continuous in both arguments. Thus, the penalty function must be continuous in α which is true by assumption. This is a weak condition which covers many penalties used in practice. The limit points may be local minima or saddle-points.

3.2. Convergence to a local or global minimum

In [5, 7], there is no simple general result for the EM that guarantees convergence to a local or global minimum without further assumptions. For instance, if the penalty function is convex, then the penalized -log-likelihood will be strictly convex and the EM algorithm will be guaranteed to converge to the unique maximum penalized likelihood value and a unique optimal image. This follows from results of [7] stated for the EM. But, if the penalty function is not convex then the sequences of the Bayesian EM estimates will only converge to a stationary point. The image at convergence will depend on the initialization of the algorithm. As noted in [5], the convergence rate (at least when the initial position is not too far from the true image) is linear and governed by the fraction of missing information, that can be evaluated from the Fisher Information matrix. Thus, here we device to use the observed part of the image Y_{obs} as an initial estimate.

Many interesting penalties that produce sparse solutions are non-convex or even non-smooth. Unfortunately, their use will be at the price of no guaranty to converge to a global or even to a local minimum. To circumvent this major drawback, we use the same heuristic argument as that of the MCA, for which we give a statistical interpretation. Indeed, one can consider the penalized -log-likelihood functional of the M-step as a Gibbs energy, where the regularization parameter parallels the temperature in the same spirit as for simulated annealing. Then, we start at a high temperature (i.e. λ) and then decrease λ according to a prescribed schedule (e.g. exponential or linear). For each value of λ, we run one iteration of the GEM inpainting Algorithm 2.4. This algorithm has flavour of a stochastic version of the EM [8].

4. EXPERIMENTAL RESULTS

The GEM inpainting algorithm was applied to several synthetic and real degraded images, from which we present few examples. Fig.1 depicts an example on Lena (input SNR=30dB). The dictionary contained the curvelet transform and the convex l_1 penalty was used. The threshold parameter was fixed to the classical value 3σ. The algorithm converged after 100 iterations.

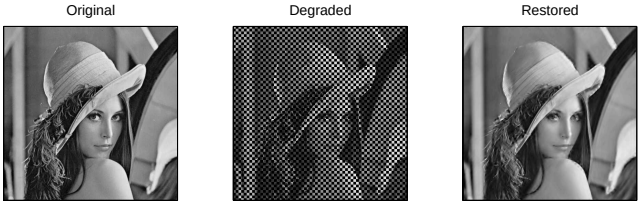

Fig. 1. Example with Lena.

To further illustrate the power of the GEM 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 old degraded photograph. The missing areas (mask of degraded parts to be recovered) were manually plotted. The result is shown in Fig.3. Here, the dictionary contained the undecimated DWT. Curvelets yield similar results.

5. CONCLUSION

We proposed a novel generalized EM-based algorithm for image inpainting. Our work is now directed towards deeper investigation of the mathematical properties (e.g. convergence behaviour and relation to stochastic versions of the EM, asymptotic properties, etc) of this algorithm.

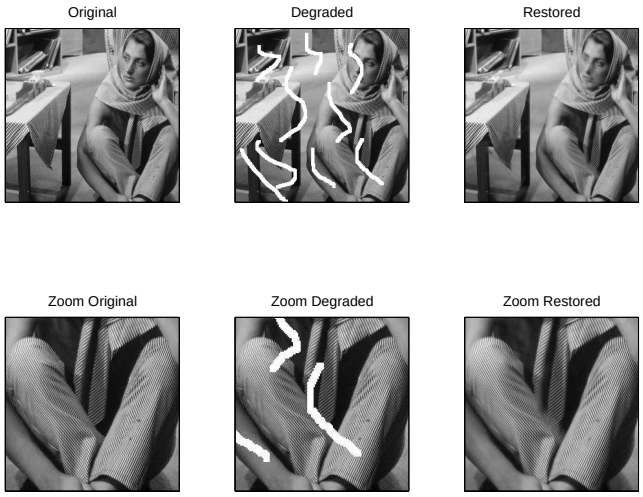

Fig. 2. Example with Barbara.

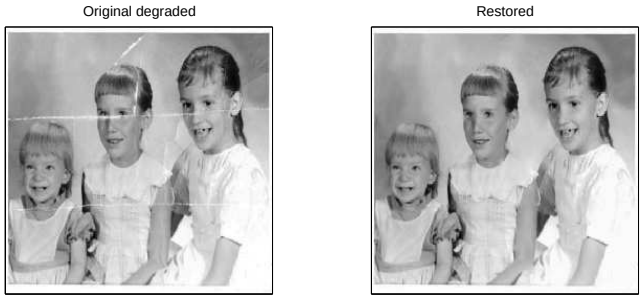

Fig. 3. GEM inpainting of a real degraded image.

6. 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," *under preparation*, 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] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc. B*, vol. 39, no. 1, pp. 1–38, 1977.

[6] R. Little and D. Rubin, *Statistical analysis with missing data*, Wiley, New York, 1987.

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

[8] G. Celeux, D. Chauveau, and J. Diebolt, "Some stochastic versions of the EM algorithm," *J. Stat. Comp. & Sim.*, vol. 55, pp. 287–314, 1996.