A Simple U-Diffusion Inpainting Structure

Jugurta Montalvão¹, Gabriel Bastos¹, and Israel Santos Filho¹

¹Affiliation not available

June 07, 2024
A Simple U-Diffusion Inpainting Structure

Jugurta Montalvão, Gabriel F. A. Bastos, and Israel J. Santos Filho,

Abstract—The inpainting problem is addressed in this work through a very simplified version of the approach based on low-dimensional manifold model (LDMM), in which the actual working principle of the LDMM is put into evidence, namely, the simulated diffusion of image pixels that takes place in a manifold from which patches are drawn to form a given image. The simplicity of this principle is translated into a straightforward algorithm that borrows ideas from the Locally Linear Embedding (LLE) method. The equivalence between this much simpler algorithm and the LDMM is illustrated through visual inspection and experimental measurements of peak signal-to-noise ratios. Additionally, a (U-shaped) multi-scale use of the proposed algorithm is presented as a significantly better initializer for missing pixels, thus reducing the number of algorithmic iterations.

Index Terms—LDMM, Multiscale, Inpainting, LLE.

I. INTRODUCTION

THE term inpainting refers to the task of automatically filling in missing or deteriorated samples from a signal or image without introducing noticeable changes from its undamaged version. Although this problem has been around for many years as a research hotspot, it has recently garnered a lot of attention due to its broad range of applications in various domains, such as object removal in information security [1], audio and image restoration [2], [3], denoising [4] and others.

Mathematically, the inpainting problem is described as the recovery of a signal $u \in \mathbb{R}^n$ from incomplete measurements
\[ x = S(u) + \eta, \tag{1} \]
where $\eta$ is the observation noise and $S(\cdot)$ stands for a linear operator that causes subsampling in the original signal $u$. Recovering $u$ from the observed $x$ is often an ill-posed problem, because even for $\eta = 0$, a given observed $x$ may come from many possible signals $u$, and some kind of regularizing \textit{a priori} is needed, such as total variation norm [5], which can eventually be replaced by nonlocal total variation in order to recover high frequency information [6]. Another class of priors are the $\ell_1$ regularizers. It was shown by Donoho [7] that the $\ell_1$ norm is a good approximation for the $\ell_0$ norm, and can thus support many techniques that impose sparsity in a given transform domain or over a learned dictionary as prior for signal reconstruction [8], [9]. Over the years, priors learned by deep convolutional neural networks in a data-driven manner have also been applied to achieve outstanding results [10], specially autoencoding models, such as the U-NET [11].

All of the aforementioned priors exploit one fundamental gear, namely, the inherent redundancy present in natural signals, which are known for having repetitive patterns and restrictions on their constructions. This redundancy may manifest in various ways, one of which is the generation of low dimensional structures, as stated by Peyré [12]. Accordingly, patches from natural images and signals tend to lie along low dimensional manifolds, in spite of their apparent high-dimensionality.

Inspired by Peyrée’s findings, a method called low dimensional manifold model (LDMM) [13] was proposed for solving inverse problems, such as inpainting. In LDMM, the dimension of the patch (chunks of a given signal/image) manifold is directly used as a regularizer. In [14], this same dimensionality criterion was interpreted as an imposition of energy concentration of convolution framelet coefficients. A more systematic exploration of this energy concentration property yielded a new improved version of LDMM, called rw-LDMM.

In this paper, we propose a simpler alternative to the LDMM by revisiting the principles stated by Peyré [12]. We show that almost equivalent results can be achieved by simple diffusion algorithms applied to the point cloud generated by signal patches. Additionally, it is also shown that by using available redundancies at different scales of observations, the number of inpainting iterations can be significantly reduced. The proposed new perspective borrows ideas from the Locally Linear Embedding (LLE) [15].

II. RESTRICTED DIFFUSION

Low dimension of patch manifolds is a key aspect for the impressive results presented in [12], [13] and [14], because it significantly increases the likelihood of finding two similar patches in a single signal instance, even if they are far apart in their independent variable domain (e.g. similar patches far apart in images, or similar sound segments far apart in time). Therefore, because similar patches are close points in their corresponding low-dimensional manifolds, for inpainting purposes, simulated diffusion through these manifolds can be used to replace missing values (e.g. missing pixels) with “diffused” near observations. On the flip side, missing values in patches also impose some kind of diffusion restriction. As an initial illustration, we borrowed from [12] the coral image, with the same inpainted set of pixels as input image, and we restored it, as shown in Fig. 1, by taking every $5 \times 5$ patch centered at an unknown pixel, comparing it to all other $5 \times 5$ patches centered at a known or previously estimated pixel, and replacing the unknown pixel with the average of central pixels of the $3$ near neighboring patches. Therefore, known pixels diffuses toward black spots, through patches of images that are neighbors in the underlying manifold.

Here, pixels are in the range from 0 (black) to 1 (white), but 10% of them are missing. Thus, to assure a proper finding of neighbors, we arbitrarily impose that a distance between two patches are to be computed only when at least 30% of

J. Montalvão was with the Department of Electrical Engineering, Federal University of Sergipe, São Cristóvão, 49000-000. (e-mail: jmontalvao@academico/ufrs.br).

Manuscript received June XX, 2024; revised XXXX XX, 20XX.
known pixels in both patches have same row and column. Furthermore, only patches at supremum distances below 0.2 (in a range from 0 to 1) are taken as near neighbors. Because of these restrictions in distance computation between patches, we refer to it as a restricted diffusion, which can be iterated, so that missing values not estimated in the first interaction are gradually approached by estimated values, until all missing values are estimated, and this diffusion of known pixels through the manifold of patches can be stopped.

As illustrated in Fig. 1, the restricted diffusion does the most relevant work regarding missing values estimation. Indeed, if we consider the visual result presented in subfigure (c’) of Figure 1 of [13], obtained with the LDMM – which is itself built upon the principles established in [12] –, this simple restricted diffusion alone, under restrictions similar to those applied in Fig. 1, yields a similar texture reconstruction, as shown in Fig. 2, including parallel details in the shawl, from 10% of known pixels, randomly taken.

Fig. 2. Illustration of how a restricted diffusion gathers non-local redundancies from known values, yielding image details reconstruction. This inpainting result is to be compared to Fig. 1 in [13].

Alas, this straightforward form of restricted diffusion yields a noisy inpainting, because similar patches all over the image are prone to small value fluctuations. To obtain a smoothed version of reconstructed images, one may take advantage of the assumed locally low dimensionality of the patch manifold and impose that near patches in the manifold are points lying on a hyperplane. To keep it as simple as possible, we use the same rational behind the LLE method, as detailed in section III.

III. PROPOSED ALGORITHM

As the restricted diffusion used to obtain Figures 1 and 2, for every patch in the image whose center is an unknown value, a set of near neighboring patches are gathered as the first algorithmic step. Here again, computing distances between patches with missing values is an important concern. Unlike the previous restricted diffusion, however, a less constrained approach is taken, where it is assumed that all patches are fulfilled with either known or previously estimated values, then distances can be computed without restrictions at this algorithmic step, while restrictions are moved to the next steps, through a dumping factor for missing values, $\Phi$. Thus, after $K_{NN}$ near neighbors of the target patch are found, their distances are compared to a threshold $\Delta_{NN}$, for a refined selection of patches. Finally, following the same idea used in the LLE method, it is assumed that the target patch can be obtained as a weighted sum of its near neighbors (because the manifold is assumed to be locally almost flat). This assumption is expressed in Eq. 2.

$$\hat{p} = C^T w,$$

where $\hat{p}$ is the estimated target patch represented as a vector, $C$ is a matrix where each row is a selected near neighbor of the actual target, $p$, and $w$ is a vector of weights to be optimized. Thus, an error vector can be defined as $e = p - \hat{p}$.

Because some elements of $e$ result from unknown values of $p$, the parameter $\Phi$ is finally used to reduce the importance of these error values, through the multiplication of $e$ by a diagonal matrix whose entries are $D(k,k) = |f(k) - \Phi|$, where $f$ is a vector of flags indicating whether a pixel is known, with flag 1, or unknown, with flag 0. For instance, if $\Phi$ is set to 0.2, error values from known pixels in $p$ become four times more relevant than those from unknown pixels.

Weight vector $w$ is then optimized to minimize the norm of the (weighted) vector of errors through pseudo-inversion, as in Eq. 3.

$$w_O = (CDC^T + \epsilon I)^{-1} CDp,$$

where $\epsilon$ is a small positive real to prevent matrix inversion singularity ($\epsilon = 10^{-4}$ in all experiments reported in this paper), and $I$ stands for the identity matrix.

The central pixel of the optimal reconstructed patch, $\hat{p} = C^T w_O$, can then be used to re-estimate the central pixel of the target patch, as detailed in Algorithm 1.

Despite its simplicity, this algorithm yields experimental results almost equivalent to those proposed in [13] and [14], which can be illustrated with the same 10% of pixels from the Barbara image, for this original subsampling was publicly shared. In Fig. 3, the performance of Algorithm 1 is illustrated, following the same initialization of unknown pixels with Gaussian random noise. After 100 iterations of the algorithm, the obtained peak signal-to-noise ratio (PSNR) was 23.32 dB, whereas more elaborated algorithms in [13] and [14] yielded 24.74 dB and 25.41 dB, respectively, for the same task. But it is noteworthy that the visual quality of the presented result is equivalent to those presented in the mentioned works (see Figure 1, in [13], and Figure 9, in [14]), which is highlighted by zoomed spots in Fig. 3.
Besides, the simplicity of this algorithm allows for a parameter selection based on some further analyses. For instance, parameter $\Phi$ can be initially set according to the proportion of known values/pixels in the image. In the former illustration with Barbara image, $\Phi$ would be set around 0.1 to compensate for the expected proportion of only 10% of known pixels in each patch. Indeed, $\Phi = 1/9$ works fine in this case, but we fine-tuned it to 0.2 in order to slightly improve the final PSNR.

### Algorithm 1 Restricted Diffusion Algorithm (single iteration)

- **Data:** $X_0$: Inpainted image, $F_0$: Flags of known values/pixels, $L$: Padding size, $\Delta_{NN}$: Maximum supremum distance, $K_{NN}$: Maximum number of nearest neighbors, $\Phi$: Weight of unknown values.
- **Maximum supremum distance, $K_{NN}$**: Maximum number of nearest neighbors, $\Phi$: Weight of unknown values.
- **Concatenate $L$ zero-valued pixels at every border of $X_0$ to increase its size;**
- **$P$** ← matrix with $M_0 \times N_0$ rows and $(2L+1)^2$ columns, where each row corresponds to a patch of $X_0$.

```
foreach $m$ and $n$ where $F_0(m, n) = 0$ (unknown value/pixel) do
    foreach $i$ and $j$ do
        $p$ ← $(2L+1)^2$-dimensional vector whose entries, $p(k)$, are values around $X_0(i, j)$;
        $f$ ← $(2L+1)^2$-dimensional vector whose entries, $f(k)$, are flags around $F_0(i, j)$;
        Find the $K_{NN}$ near neighbors of $p$ among the rows of $P$;
        $r_c$ ← row index of the $c$-th collected near neighbor ($c \in \{1, 2, \ldots, C\}$, $C \leq K_{NN}$) whose distance is $\leq \Delta_{NN}$;
        if Number of collected row indices $> 0$ then
            $D$ ← diagonal matrix whose entries are $D(k, k) = |f(k) - \Phi|$;
            $C$ ← matrix whose $C$ rows are taken from $P$;
            $p$ ← $C^T(D^T + \epsilon I)^{-1}CDp$;
            $X_0(i, j)$ ← $\frac{1}{2}(X_0(i, j) + p(\text{center}))$;
        end
    end
end
```

Fig. 3. Performance illustration for the proposed inpainting algorithm, where 90% of pixels are missing (left). These missing pixels are initialized with noise (center) and given as input to the algorithm, whose output (right) has a PSNR of 23.32 dB. Reconstruction details are zoomed in for better visual comparisons.

To set parameter $\Delta_{NN}$, one may use the intrinsic dimensionality analysis, as explained in [16], based on the Grassberger-Proccacia algorithm. Accordingly, a multiscale analysis can be done, as shown in Fig. 4, with a subsample of patches randomly drawn from the image. For this example, 5000 patches ($17 \times 17$) were randomly drawn from the image, and two relevant estimates are shown in Fig. 4. These two relevant estimates suggest that patches gathered in hypercubes of edge around $\Delta = 2^{-2.5}$ tend to be on a flat surface whose local dimension is between 2 and 3. Indeed, a local dimension around 2D is expected due to the superposition of patches, as explained in [12]. The second estimate, in a larger scale of analysis, around $\Delta = 2^{-1}$, seems to be a more interesting one, for inpainting purposes, for it suggests that patches gathered in hypercubes of edge around $\Delta = 1/2$ tend to be on a flat 5-D surface, which seems to be the consequence of image texture redundancies, beyond the 2 degrees of freedom due to patch superposition. Therefore, we may initially set $\leq \Delta_{NN}$ to 0.5 to a reasonably inpainting. Indeed, from this initial value, we fine-tuned it to 0.7, for a better result in terms of PSNR. The flip side of this approach is that, because the perfect image is known only in controlled experiments, but not in real applications of image recovery from their sparse representations, the Grassberger-Proccacia should be applied to similar images, when available. Otherwise, $\leq \Delta_{NN}$ must be arbitrarily set in the range from 0 to 1.

### A. U-shaped restricted diffusion

Initialization of unknown pixels with gaussian noise, with the same mean and variance of the known pixels, is a usual step for preparing input images in LDMM and rw-LDMM, but a multiscale approach can replace it with advantages, where both the initial subsampled image and matrix of flags – level 0 – are reduced through a maxpooling downsampling until the number of unknown pixels is null – level B. Then the inpainting algorithm is applied to the image at level B+1. After stabilization of the diffusion process at this level, the restored image is upsampled with a transposed convolution to level B+2, where the diffusion process takes place again, and this is repeated until level 0 is attained, as illustrated in Fig. 5.

Fig. 6 provides a perspective of the much faster evolution of the U-shaped approach toward its final performance.
Fig. 5. Illustration of the U-shaped structure of multiscale diffusion.

Fig. 6. Comparison between diffusion stabilization through iterations. The U-shaped approach provides a much better initialization at level 0 than initialization with noise. It also provides a slightly better final PSNR.

IV. EXPERIMENTS

Inpainting experiments were done with images also used in [13] and [14], to allow direct comparisons in terms of PSNR, as in Table I.

<table>
<thead>
<tr>
<th>Image</th>
<th>LDMM</th>
<th>rw-LDMM</th>
<th>U-Diffusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>24.74–24.78</td>
<td>25.34–25.59</td>
<td>23.98</td>
</tr>
<tr>
<td>Boat</td>
<td>23.08–23.21</td>
<td>23.31–23.66</td>
<td>23.21</td>
</tr>
<tr>
<td>Couple</td>
<td>23.04–23.78</td>
<td>23.62–24.27</td>
<td>23.07</td>
</tr>
<tr>
<td>Checkerboard</td>
<td>12.18–12.37</td>
<td>13.74–14.29</td>
<td>12.26</td>
</tr>
</tbody>
</table>

The term U-Diffusion refers to the proposed algorithm, including the U-shaped chain of multiscale diffusions. Results for the LDMM and rw-LDMM were taken from the references, whereas results from the proposed U-Diffusion were not optimized. Those results suggest that, although PSNR values are in the same range, the U-Diffusion remains below results obtained with the rw-LDMM. However, as illustrated in Fig. 3, in all experiments, recovered images are qualitatively alike, as compared to images inpainted with competing methods.

Besides its conceptual simplicity, another important advantage of the proposed algorithm is its reduced computational cost. To inpaint an \( N \times N \) image, the proposed algorithm requires at most \( N^2 \) pseudo-inversions of \( \ell \times K_{NN} \) matrices at every iteration (for the number of near neighbors can be less than \( K_{NN} \)), where \( \ell \) is the patch size, which has a computational complexity of \( \mathcal{O}(N^2 \ell K_{NN}^2) \), in addition to nearest neighbors searches, which are common for all algorithms. In contrast, every iteration of the LDMM requires solving \( \ell \times N^2 \times N^2 \) linear systems. Using GMRES [19], the time complexity of it is \( \mathcal{O}(\ell k \text{nnz}) \), where \( k \) is the number of GMRES iterations and \( \text{nnz} \) is the number of nonzero entries in the weighted adjacency matrix, which is usually \( 50N^2 \), since only the 50 nearest neighbors are considered to make the matrix sparse. Considering the typical values of the parameters mentioned in both papers, and a reasonable amount of GMRES iterations (\( k = 50 \)), we estimate that, without considering nearest neighbor queries, an iteration of our method can be about 2 to 3 times faster than a LDMM one. As for rw-LDMM, the cost is even higher, since \( \ell + r \) systems are solved, and calculating the partial singular value decomposition of a \( N^2 \times \ell \) matrix is also required.

V. CONCLUSION

A new inpainting algorithm was proposed, which is built upon the seminal ideas found in [12], but also guided by a search for simplicity. We gathered experimental evidences that the very simple principle of diffusion through a manifold of patches, under some restrictions imposed by missing values, is the main gear behind the inpainting results obtained with the LDMM and the rw-LDMM, both presented in a convoluted articulation of abstract concepts. In this work, it was shown that a much simpler perspective (and implementation) is enough to yield similar results, in terms of PSNR, with virtually indistinguishable visual results.

Besides, a U-shaped structure for the algorithm was proposed, in which a multiscale handling of subsampled images yields a much better starting point for algorithmic iterations at the highest resolution scale, many dB of PSNR above usual initialization with noise. This initial advantage can be used to significantly reduce computational burden. Besides, by comparing the computational burden of each iteration, the proposed algorithm shows to be less time consuming, as compared to both the LDMM and the rw-LDMM.

Given the myriad of potential applications of a method such as the LDMM, which is able to solve other ill-posed problems, besides inpainting – e.g. denoising, super resolution –, we believe that the simplicity of the method proposed in this work, including its multiscale U-shaped structure, may be a useful tool for researchers and practitioners with many different theoretical backgrounds, for it can be easily implemented and modified for further experimentation and development of new related ideas.
REFERENCES