ANALYSIS OF A VARIATIONAL FRAMEWORK FOR EXEMPLAR-BASED IMAGE INPAINTING

P. ARIAS, V. CASELLES, G. FACCIOLO*

Abstract. In this paper we study some variational models for exemplar-based image inpainting, namely the patch non-local means and the patch non-local Poisson methods, previously studied by the authors from the experimental point of view. In both cases, the unknowns are the image u to be reconstructed and a weight function w expressing the similarity of patches. As a limit case of the studied framework, the weight function reduces to a correspondence map from the inpainting domain to the know part of the image. We prove the existence and regularity of minima for both functionals. In particular, we prove the existence of optimal correspondence maps which are uniform limits of maps of bounded variation with finitely many values. Then we prove the convergence of an alternating optimization scheme for the variables (u, w). We also prove the convergence in probability of the PatchMatch method, a recently introduced and efficient algorithm to compute optimal correspondence maps. Finally, we display some numerical experiments illustrating the performance and properties of the methods.

Key words. image inpainting, variational models, exemplar-based, non-local means

AMS subject classifications. 68U10, 35A15, 65D05, 65C50

1. Introduction. The purpose of the present paper is the analysis of some variational models for non-local image inpainting. *Image inpainting*, also known as *image completion* or *disocclusion*, aims to obtain a visually plausible image interpolation in a region of the image in which data are missing due to damage or occlusion. It has become a standard tool in digital photography for image retouching (e.g. the removal of scratches in old photographies) and it is the object of intensive research in order to convert it into a key tool for video post-production. Besides its numerous applications, the problem is of theoretical interest since its analysis involves an understanding of the self-similarity present in natural images, visible in repetitive geometric and texture patterns that appear in almost any image.

Most inpainting methods found in the literature can be classified into two groups: *geometry-* and *texture-oriented* methods. We now briefly review the developments in both types of approaches, with emphasis in texture-oriented methods. This review will be helpful for motivating the proposed formulation.

Geometry-oriented methods. In this class of methods images are usually modeled as functions with some degree of smoothness, expressed for instance in terms of the curvature of the level lines or the total variation of the image. They take advantage of this smoothness assumption and interpolate the inpainting domain by continuing the geometric structure of the image (its level lines, or its edges), usually as the solution of a (geometric) variational problem or by means of a partial differential equation (PDE). Such PDE can be derived from variational principles, as for instance in [44, 8, 19, 20, 27, 43], or inspired by phenomenological modeling [11, 15, 54]. These methods show good performance in propagating smooth level lines or gradients, but fail in the presence of texture. They are often referred to as *structure* or *cartoon* inpainting.

In most cases, geometry-oriented methods are *local* in the sense that they are based on PDEs. An implication of this is that among all the data available from the image, they only use that at the boundary of the inpainting domain.

*DTIC, Universitat Pompeu-Fabra, Barcelona, Spain,
e-mail: {pablo.arias; gabriele.facciolo; vicent.caselles}@upf.edu

Texture-oriented methods. Texture-oriented inpainting was born as an application of texture synthesis, *e.g.* [25, 33]. Its recent development was triggered in part by the works of Efros and Leung [25] and Wei and Levoy [55] using non-parametric sampling techniques (parametric models have also been considered, *e.g.* [41]). In these works texture is modeled as a two dimensional probabilistic *graphical model*, in which the value of each pixel is conditioned by its neighborhood. These approaches rely directly on a sample of the desired texture to perform the synthesis. The value of each target pixel x is copied from the center of a (square) patch in the sample image, chosen to match the available portion of the patch centered at x. See [42] for a probabilistic theoretical justification.

This strategy (with various modifications) has been extensively used for inpainting [12, 14, 21, 23, 25, 45]. As opposed to geometry-oriented inpainting, these so-called *exemplar-based* approaches, are *non-local*: to determine the value at x, the whole image may be scanned in the search for a matching patch.

As pointed out in [22] the problem of exemplar-based inpainting can be stated as that of finding a *correspondence map* $\varphi : O \to O^c$, which assigns to each point x in the inpainting domain O (a subset of the image domain Ω, usually a rectangle in $\mathbb{R}^2$) a corresponding point $\varphi(x) \in O^c := \Omega \setminus O$ where the image is known (see Figure 1.1). The unknown part of the image is then synthesized using the map φ. The filling-in strategy of [25, 55] can be regarded as a *greedy* procedure (each hole pixel is visited only once) for computing a *correspondence map*. The results obtained are very sensitive to the order in which the pixels are processed [21, 45, 32].

To address this issue, in [22] the authors proposed to model the inpainting problem as the minimization of an energy functional in which the unknown is the correspondence map itself:

$$E(\varphi) = \int_O \int_{\Omega_p} |\hat{u}(\varphi(x+h)) - \hat{u}(\varphi(x)+h)|^2 \mathrm{d}h\mathrm{d}x, \tag{1.1}$$

where $\hat{u} : O^c \to \mathbb{R}$ is the known part of the image, and Ω_p is the patch domain (a neighborhood of the origin $0 \in \mathbb{R}^2$). The unknown image is computed as $u(x) = \hat{u}(\varphi(x))$, $x \in O$. Thus φ should map a pixel x and its neighborhood in such a way that the resulting patch is close to the one centered at $\varphi(x)$. This model has been the subject of further analysis by Aujol *et al.* [6], proposing extensions and proving the existence of a solution in the set of piecewise roto-translation maps, *i.e.* maps of the form

$$\varphi(x) = \sum_{i \in I} R_i(x - c_i)\mathbb{1}_{A_i}(x),$$

where $\{A_i\}_{i \in I}$ is a Caccioppoli partition of O (i.e. all sets of the partition have finite perimeter in O and the sum of the perimeters is finite), and for each $i \in I$ R_i is a rotation matrix and c_i is a translation vector. $\mathbb{1}_{A_i}(x) = 1$ if $x \in A_i$ and zero otherwise.

The energy (1.1) is highly non-convex and no effective way to minimize it is known [6]. Hence, other authors have addressed the determination of a correspondence map by looking for simpler optimization problems. In [39] the problem is formulated as probabilistic inference on a graphical model. Using a message passing algorithm the authors efficiently compute a coarse correspondence map.

Another optimization strategy is followed in [37, 57]. In both works the variable to be optimized is the unknown image whereas the correspondence map appears as

an auxiliary variable. The resulting energy functional can be regarded as a relaxation of (1.1):

$$E(u,\varphi) = \int_{\widetilde{O}} \int_{\Omega_p} |u(x+h) - \hat{u}(\varphi(x)+h)|^2 \mathrm{d}h\mathrm{d}x, \tag{1.2}$$

where $\widetilde{O} := O + \Omega_p$ denotes the set of centers of patches that intersect the inpainting domain O (see Figure 1.1). The energy is usually optimized using an alternating scheme with respect to the variables u and φ, and the unknown image is determined as part of the optimization process, and is not constrained to be $u(x) = \hat{u}(\varphi(x))$. Although this relaxation is still non-convex, the alternating minimization scheme converges to a critical point of the energy. This approach was also used in the context of texture synthesis [40].

Exemplar-based methods provide impressive results in recovering textures and repetitive structures. However, their ability to recreate the geometry without any example is limited and not well understood. Different strategies have been proposed for combining geometry and texture inpainting. Some rely on human intervention for constraining the geometry [53]. Others usually decompose the image in structure and texture components. The structure is reconstructed using some geometry-oriented scheme, and this is used to guide the texture inpainting [12, 18, 23, 35].

Let us finally note that the works in texture synthesis of [25, 55] have also influenced the development of non-local methods for other applications, such as denoising [7, 17], superresolution [50] and regularization of inverse problems [30, 48]. As opposed to the case of inpainting, in these contexts the denoised pixel value is estimated from many locations in the image, typically as a (non-local) average. This results in replacing the correspondence map by a weight function $w : \Omega \times \Omega \to \mathbb{R}$, with Ω being the image domain (usually a rectangle in $\mathbb{R}^2$). For each x, $w(x,\cdot)$ weights the contribution of each image location to the estimation of x. Inspired by regularization techniques used in the context of graphs or discrete data and trying to formulate the non-local means denoising method [17, 7] as a variational model, Gilboa and Osher [30, 31] proposed the following functional

$$E_w(u) = \int_\Omega \int_\Omega w(x,y)(u(x) - u(y))^2 \mathrm{d}y\mathrm{d}x \tag{1.3}$$

which can be considered as a non-local version of the Dirichlet integral. The weights w are considered as known. The minimum of (1.3) should have a low pixel error $(u(x) - \hat{u}(y))^2$ whenever $w(x,y)$ is high. In this way the weights drive the information transfer from known to unknown pixels. When the weights are *Gaussian*

$$w(x,y) \propto \exp\left(-\frac{1}{T}\|p_u(x) - p_{\hat{u}}(y)\|^2\right), \tag{1.4}$$

the non-local means algorithm results from the first step of Jacobi's iterative method for solving the Euler-Lagrange equation of (1.3). Here, $\|\cdot\|$ is a weighted L^2-norm in the space of patches and T is a parameter that determines the selectivity of the weights w.

Other variational approaches for non-local denoising have also been proposed in [38, 16, 48, 49].

1.1. Our contribution and the organization of the paper. In our previous work [4, 5] we proposed a variational framework for image inpainting (see Section 2), and discussed some of its properties, mainly from a practical point of view. The present work complements [5] with a theoretical study of some aspects of the framework. Before enumerating the contributions, let us briefly point out the main characteristics of the framework proposed in [5].

Our variational formulation is inspired by (1.3). In [31] the weights are considered as known and remain fixed through all the iterations. While this might be appropriate in applications where they can be estimated from the degraded image, in the image inpainting scenario here addressed, they are not available and have to be inferred together with the image (see also [7, 47, 50]).

As in (1.3) we encode the image redundancy as a non-local weight function $w(x, y)$ measuring the similarity of the patch around $x \in \widetilde{O}$ with the patch around $y \in \widetilde{O}^c$, where $\widetilde{O}^c$ denotes the complement of $\widetilde{O}$ (we refer to Section 2 for precise definitions). One of the novelties of the models proposed in [5], is the inclusion of adaptive weights in a variational setting, considering the weight function w as an additional unknown. Instead of prescribing explicitly the Gaussian functional dependence of w w.r.t. u, we do it implicitly, as a component of the optimization process.

On one hand this permits to write a simpler functional which has the form of a Gibbs free energy with temperature parameter $T > 0$, or its limit case where $T = 0$. On the other, this is in consonance with the alternating optimization algorithm, one of the algorithms used in practice.

Let us now describe the contributions of this work.

Connections with statistical mechanics and probabilistical inference. In Section 3 we discuss the connections of our model with similar functionals appearing in statistical mechanics. These connections help to understand the formulation by drawing analogies with existing and well known theories. Let us mention in particular the approach to quantization and clustering in terms of an statistical mechanics formulation (see [51] and references therein).

Mathematical analysis of variational models for inpainting. The framework proposed in [5] allows the derivation of different inpainting models. In [5] we discussed four of them. The present work complements [5] with a theoretical study of two models contained in the general Gibbs energy formulation: the patch NL-means, based on a patch-based formulation of (1.3), and the patch NL-Poisson model which replaces the image in the patch NL-means energy by its gradients. For both of them we prove in Section 4 the existence of minima and their regularity.

In our formulation, we constrain $w(x, \cdot)$ to be a probability density function, which can be seen as a relaxation of the correspondence maps used in [6, 22], providing a fuzzy correspondence between each $x \in O$ and the points in the complement of the inpainting domain.

Analysis of the correspondence model. In Section 5 we study the Gamma limit as $T \to 0$ of the patch NL-means energy. This results in a relaxation of the energy (1.2). We prove the existence of optimal solutions of the limit functional which are measurable correspondence maps which also minimize (1.2) (a model computationally studied in [57, 4, 37, 5]).

Heuristically, one can say that the optimal similarity weights w converge to $\delta(y - \varphi(x))$ where $\varphi : \widetilde{O} \to \widetilde{O}^c$ is an optimal correspondence map for (1.2). This case is the most relevant for the image inpainting application. Of particular interest, is the regularity of the correspondence map obtained by minimizing (1.2). We prove a mild

regularity result, namely the existence of optimal correspondence maps φ which are uniform limits of maps of bounded variation with finitely many values. This result is interesting in connection with the experimental observation that the computed φ copies rigidly parts of the image outside the inpainting domain, behaving locally as a translation (see Fig. 1.1). This observation is also at the root of the roto-translation constraint imposed to φ in [6].

Convergence of the alternating optimization scheme. In Section 6 we prove the convergence to critical points of the alternating optimization scheme with respect to the variables u and w (which coincides with the Expectation-Maximization (EM) algorithm) for both models, patch NL-means and -Poisson.

Discretization: mathematical analysis of PatchMatch [9] algorithm. In Section 7 we study an efficient algorithm to minimize (1.2), called the PatchMatch [9]. The most time consuming step in the minimization of (1.2) is the computation of the optimal matching between patches in the inpainting domain and patches in the region of available data. Recently Barnes *et al.* [9] introduced the PatchMatch, an efficient algorithm based on heuristics to solve the problem of matching patches between images. In this paper we prove its convergence in probability and we compute a bound on its convergence rate.

Finally, in Section 8 we display some experiments to illustrate the main features of the algorithms and the results obtained with the proposed models.

Figure 1.1: **Inpainting problem.** Left: on a rectangular image domain Ω, missing data u in a region O has to be reconstructed using the available image $\hat{u}$ over $O^c := \Omega \setminus O$. A patch centered at $x \in O$ is denoted by $p_u(x)$. The set of centers of incomplete patches is $\widetilde{O} := O + \Omega_p$, where Ω_p denotes the patch domain. Middle: the image shows a completion obtained using the *patch NL-medians*, a scheme derived from the general formulation presented in this work. Right: the resulting completion shows a correspondence map which is a piece-wise translation. The red curves show the boundaries between the regions of constant translation. In each of these regions, the image is copied rigidly from a corresponding region in O^c.

2. A variational framework. In this Section we review the variational framework for non-local image inpainting proposed in [5], which is inspired by recent developments on image regularization and denoising [17, 38, 31].

Let us introduce some notation. Images are denoted as functions $u : \Omega \to \mathbb{R}$, where Ω denotes the image domain, usually a rectangle in $\mathbb{R}^N$. We will commonly refer to points in Ω as pixels. These will be denoted by x, y or h, the latter for positions inside the patch. A patch of u centered at x is denoted by $p_u(x) = p_u(x, \cdot) : \Omega_p \to \mathbb{R}$

where Ω_p is a rectangle centered at 0. The patch is defined by $p_u(x,h) = u(x+h)$, with $h \in \Omega_p$. Let $O \subset \Omega$ be the hole or inpainting domain, and $O^c = \Omega \setminus O$. We assume that O is an open set with Lipschitz boundary. We still denote by u the part of the image u inside the hole, while $\hat{u}$ is the part of u in O^c: $\hat{u} = u|_{O^c}$.

Let us define some domains in Ω that are useful to work with patches. We denote by $\widetilde{\Omega}$ the set of centers of patches contained in the image domain, i.e. $\widetilde{\Omega} = \{x \in \Omega : x + \Omega_p \subseteq \Omega\}$. As was defined in the Introduction, we take $\widetilde{O}$ as the set of centers of patches that intersect the hole, *i.e.* $\widetilde{O} = O + \Omega_p = \{x \in \Omega : (x + \Omega_p) \cap O \neq \emptyset\}$. For a simplified presentation, we assume that $\widetilde{O} \subseteq \widetilde{\Omega}$, *i.e.* every pixel in $\widetilde{O}$ is the center of a patch contained in Ω. We denote $\widetilde{O}^c = \widetilde{\Omega} \setminus \widetilde{O}$. Thus, patches $p_{\hat{u}}(y)$ centered at points $y \in \widetilde{O}^c$ are contained in O^c (see Figure 1.1). Further notation will be introduced in the text.

2.1. Review of the formulation. In this setting, we consider an energy which contains two terms, one of them is inspired by (1.3) and measures the coherence between patches in $\widetilde{O}$ and those in $\widetilde{O}^c$, given a similarity weight function $w : \widetilde{O} \times \widetilde{O}^c \to \mathbb{R}$. This permits the estimation of the image u when the weights w are known. The second term allows us to compute the weights given the image. Thus, the complete functional is

$$\begin{aligned} &\mathcal{E}_{\varepsilon,T}(u,w) = \mathcal{U}_\varepsilon(u,w) - T\int_{\widetilde{O}} \mathcal{H}(w(x,\cdot))\mathrm{d}x, \\ &\text{subject to } \int_{\widetilde{O}^c} w(x,y)\mathrm{d}y = 1, \end{aligned} \tag{2.1}$$

where

$$\mathcal{U}_\varepsilon(u,w) = \int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)\varepsilon(p_u(x) - p_{\hat{u}}(y))\mathrm{d}y\mathrm{d}x, \tag{2.2}$$

$\varepsilon(\cdot)$ is an error function for image patches (such as the squared L^2-norm), and

$$\mathcal{H}(w(x,\cdot)) = -\int_{\widetilde{O}^c} w(x,y)\log w(x,y)\mathrm{d}y$$

is the entropy of the probability $w(x,\cdot)$, $x \in \widetilde{O}$.

Let us now make some additional comments on the functional. We observe that the term $(u(x) - \hat{u}(y))^2$ in (1.3), that penalizes differences between pixels, is substituted in (2.2) by the patch error function $\varepsilon(p_u(x) - p_{\hat{u}}(y))$. This has several implications. First, observe that minimizing (2.2) with respect to the image u will force the patch $p_u(x)$ to be similar to $p_u(y)$ whenever $w(x,y)$ is high. Second, we observe that for a given completion u, and for each $x \in \widetilde{O}$, the optimum weights minimize the mean patch error for $p_u(x)$, given by

$$\int_{\widetilde{O}^c} w(x,y)\varepsilon(p_u(x) - p_{\hat{u}}(y))\mathrm{d}y,$$

while maximizing the entropy. This can be related to the *principle of maximum entropy* [34], widely used for inference of probability distributions. According to it, the best representation for a distribution given a set of samples is the one that maximizes the entropy, *i.e.* the distribution which makes less assumptions about the process. Taking ε as the squared L^2-norm of the patch, then the resulting weights

are given by formula (1.4). The parameter T in (2.1) controls the trade-off between both terms and is also the selectivity parameter of the Gaussian weights. Note also that by restricting $w(x,\cdot)$ to be a probability, trivial minima of E with $w(x,y)=0$ everywhere are discarded.

Let us describe in detail the main ingredients of the model and give examples of patch error functions of interest in practice.

2.1.1. The patch error function ε. Patches are functions defined on Ω_p, and if $\mathbb{P}$ denotes a suitable space of patches, we consider error functions $\varepsilon : \mathbb{P} \to \mathbb{R}^+$ defined either as the weighted sum of pixel-wise errors

$$\varepsilon(p_u(x) - p_{\hat{u}}(y)) = g * \mathrm{e}(u(x+\cdot) - \hat{u}(y+\cdot)).$$

where $\mathrm{e} : \mathbb{R} \to \mathbb{R}^+$, or gradient errors

$$\varepsilon(p_u(x) - p_{\hat{u}}(y)) = g * \mathrm{e}(\nabla u(x+\cdot) - \nabla \hat{u}(y+\cdot)).$$

where $\mathrm{e} : \mathbb{R}^2 \to \mathbb{R}^+$. Here, $g : \mathbb{R}^N \to \mathbb{R}^+$ denotes a suitable *intra-patch* kernel function. As an example, one could take $g(h) = \mathcal{N}(h|0, a\boldsymbol{I})$, the bivariate Gaussian probability density function with 0 mean and isotropic standard deviation a. In our mathematical statements we will consider a function g with compact support in $\mathbb{R}^N$. Let us give some examples of patch error functions.

Patch NL-means. In this case we use $\mathrm{e}(r) = |r|^2$ and the patch error function $\varepsilon(p_u(x) - p_{\hat{u}}(y))$ is a weighted squared L^2-norm that we denote by

$$\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,2} = g * |u(x+\cdot) - \hat{u}(y+\cdot)|^2.$$

$\mathbb{P}$ can be taken as the set of L^2 functions in Ω_p.

Patch NL-medians. In this case we set $\mathrm{e}(r) = |r|$ and we define the patch error function as a weighted L^1-norm. $\mathbb{P}$ can be taken as the set of L^1 functions in Ω_p.

Patch NL-Poisson. Let us take $\mathbb{P}$ as the space $W^{1,2}(\Omega_p)$. We consider the patch error function

$$\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,\nabla} = g * |\nabla u(x+\cdot) - \nabla \hat{u}(y+\cdot)|^2.$$

Patch NL-gradient medians. We take $\mathbb{P}$ as the space of bounded variation functions in Ω_p [3] and consider the patch error function $g * |\nabla u(x+\cdot) - \nabla \hat{u}(y+\cdot)|$, where in an abuse of notation we have used ∇ to denote the distributional derivative.

The last two patch error functions are based on the gradient of the image. As it will be discussed below, the patch error function determines not only the similarity criterion but also the image synthesis, and thus is a key element in the framework. Let us mention that the use of nonlocal energies with gradient terms for deblurring and denoising problems has been proposed in [38].

In the present work we will focus on the energies corresponding to the patch NL-means and -Poisson error functions (and their combination). The models based on the patch NL-medians and -gradient medians are discussed in [5], mainly from a practical point of view.

2.1.2. The Euler-Lagrange equations. Let us compute the Euler-Lagrange equations of $\mathcal{E}_{\varepsilon,T}$ with respect to both the weights and the image.

If we keep u fixed and we minimize (2.1) with respect to w, writing the Euler-Lagrange equation $\delta_w \mathcal{E}_{\varepsilon,T}(u,w) = 0$ we obtain

$$w_{\varepsilon,T}(u;x,y) := \frac{1}{Z_{\varepsilon,T}(u;x)} \exp\left(-\frac{1}{T}\varepsilon(p_u(x) - p_{\hat{u}}(y))\right),$$

where the normalizing factor $Z_{\varepsilon,T}(u;x)$ is given by

$$Z_{\varepsilon,T}(u;x) := \int_{\widetilde{O}^c} \exp\left(-\frac{1}{T}\varepsilon(p_u(x) - p_{\hat{u}}(y))\right) \mathrm{d}y. \tag{2.3}$$

The weight function $w_{\varepsilon,T}(x,y)$ measures the similarity between the patches centered at $x \in \widetilde{O}$ and $y \in \widetilde{O}^c$. It can be interpreted as a relaxation of an one-to-one correspondence map, establishing a fuzzy correspondence between each $x \in O$ and the complement of the inpainting domain.

For computing the Euler-Lagrange equation with respect to the image, we will consider the energies corresponding to the patch NL-means and -Poisson patch error functions. Although at this moment, these computations are formal, they help in understanding the model. Their justification follows by the results in Section 4.

Patch NL-means. The resulting Euler-Lagrange equation is the following:

$$u(z) = \frac{1}{k(w;z)} \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z - \cdot, z' - \cdot))\hat{u}(z')\mathrm{d}z', \quad z \in O,$$

where

$$w_{\widetilde{O},\widetilde{O}^c}(z,z') = \begin{cases} w(z,z') & \text{if } (z,z') \in \widetilde{O} \times \widetilde{O}^c, \\ 0 & \text{if not,} \end{cases}$$

$$g * (w_{\widetilde{O},\widetilde{O}^c}(z - \cdot, z' - \cdot)) := \int_{\mathbb{R}^N} g(h)\chi_{\widetilde{O}^c}(z' - h)\chi_{\widetilde{O}}(z - h)w(z - h, z' - h)\,\mathrm{d}h, \tag{2.4}$$

and

$$k(w;z) := \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z - \cdot, z' - \cdot))\mathrm{d}z' = 1, \tag{2.5}$$

assuming that both the weights and g are normalized. Thus, optimal u are given by a non-local average of the known pixels. The weights in the average are obtained by convolving the Gaussian similarity weights with the patch kernel g (as in [38, 49]).

Patch NL-Poisson. In this case we have that u is a solution of the Poisson equation:

$$\begin{cases} \Delta u(z) = \operatorname{div} \boldsymbol{v}(w;z), & z \in O, \\ u = \hat{u} & \text{in } \partial O, \end{cases} \tag{2.6}$$

where

$$\boldsymbol{v}(w;z) = \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z - \cdot, z' - \cdot))\nabla\hat{u}(z')\mathrm{d}z'. \tag{2.7}$$

We used again the notation (2.4) and the fact that $k(w;z) = \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z - \cdot, z' - \cdot))\mathrm{d}z' = 1$ for all $z \in O$.

We observe that the solutions this Poisson equation are minimizers of the functional $\int_{\widetilde{O}} \|\nabla u(z) - \boldsymbol{v}(w;z)\|_2^2 \,\mathrm{d}z$. Therefore, u is computed as the image with the closest gradient (in the L_2 sense) to a *guiding* vector field $\boldsymbol{v}(w;z)$ computed as a non-local average of the image gradients in the known portion of the image.

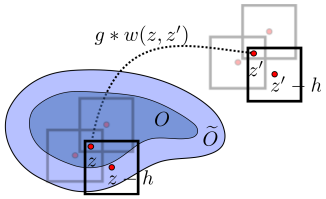

Figure 2.1: Patch-wise non-local means inpainting. The value at $z \in O$ is computed using all the patches that overlap z. The patch centered at $z - h$, contributes with the terms $w(z - h, z' - h)\hat{u}(z')$ to the average.

2.1.3. Getting a correspondence. Finally, let us note that we can get formally a correspondence map by taking the limit $T \to 0$. The resulting energy is dominated by the image term and can be written as

$$E(u, w) \simeq \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x, y)\varepsilon(p_u(x) - p_{\tilde{u}}(y))\mathrm{d}y\mathrm{d}x. \tag{2.8}$$

We will describe more precisely this limit process in Section 5 and show that the map $x \to w(x, \cdot)$ has to be replaced by a Young measure (a general measurable probability valued map). In that context there are minima of (2.8) w.r.t. the weights $w(x, \cdot)$ which can be written as Dirac's delta function on a point $\varphi(x)$ which is a nearest neighbor of the patch $p_u(x)$ with respect to the patch error function in (5), *i.e.* $w(x, y) = \delta(y - \varphi(x))$. We notice that the set $\arg\min_{y \in \widetilde{O}^c} \varepsilon(p_u(x) - p_{\hat{u}}(y))$ is not necessary made of a single point. In case that is a singleton, the problem of minimizing the energy (2.8) can be rewritten in terms of the correspondence map

$$E(u, \varphi) \simeq \int_{\widetilde{O}} \varepsilon(p_u(x) - p_{\tilde{u}}(\varphi(x)))\mathrm{d}x. \tag{2.9}$$

The model of [37, 57] (equation (1.2)) is obtained as a particular case when ε is the squared L^2-norm. An equivalent formulation has been proposed by Peyré [46], where the energy is interpreted as a regularization model based on the distance to the manifold of known patches.

Although the case $T = 0$ is the most relevant for the image inpainting application [4, 5] and most of the experiments shown correspond to $T = 0$, we will present the general framework with $T \geq 0$. This will give us a broader view of the model and the main ideas underlying it, allowing us to relate it with other models recently proposed for non-local image regularization. In this way, many of the arguments and ideas exposed next for the context of image inpainting, may be applied as well to other contexts (see in particular [28] where a variant of the formalism is applied to the interpolation of sparsely sampled images). The general context permits also to make explicit the connections with the Gibbs energy functional in statistical mechanics [26] and points towards the literature on clustering and classification formulated in that context [51]. Let us describe this in next Section.

3. Connections with statistical mechanics. Statistical mechanics has also been the framework for a variational approach to many problems like clustering, pattern recognition and classification, regression, or coding theory (see [51] and references therein). Although we will not give a detailed review of it, let us mention the basic

framework. For simplicity, we restrict ourselves to the discrete case where O is a subset of a domain Ω in $\mathbb{Z}^2$.

In the context of statistical mechanics we consider that for each $x \in \widetilde{O}$ there is a set of possible configurations indexed by the parameter $y \in \widetilde{O}^c$ with a probability density $w(x,y)$. We follow the tradition in statistical mechanics and use the notation $\beta = \frac{1}{T}$. If we consider that each configuration has an energy

$$U_\varepsilon(u;x,y) := \varepsilon(p_u(x) - p_{\hat{u}}(y)),$$

then the Boltzmann distribution gives the probability that the "particle" x is in the state y. This distribution is given by

$$p_{\varepsilon,\beta}(u;x,y) = \frac{e^{-\beta U_\varepsilon(u;x,y)}}{Z_{\varepsilon,\beta}(u;x)}, \tag{3.1}$$

where $Z_{\varepsilon,\beta}(u;x)$ is the normalization factor (also referred to as partition function).

The energy $\mathcal{E}_{\varepsilon,\beta}(u,w)$ can be written as

$$\mathcal{E}_{\varepsilon,\beta}(u,w) = \beta^{-1} \sum_{x\in\widetilde{O}} \mathrm{KL}(w(x\,\cdot), p_{\varepsilon,\beta}(u;x,\cdot)) + \sum_{x\in\widetilde{O}} F_{\varepsilon,\beta}(u;x), \tag{3.2}$$

where $\mathrm{KL}(P,Q)$ denotes the Kullback-Leibler divergence between two probability distributions P and Q in $\widetilde{O}^c$: $\mathrm{KL}(P,Q) = \sum_{y\in\widetilde{O}^c} P(y) \log\left(\frac{P(y)}{Q(y)}\right)$, and $F_{\varepsilon,\beta}(u;x)$ denotes the Helmholtz free energy $F_{\varepsilon,\beta}(u;x) := -\frac{1}{\beta}\log Z_{\varepsilon,\beta}(u;x)$. The expression

$$G_{\varepsilon,\beta}(u,w;x) := \frac{1}{\beta}\mathrm{KL}(w(x\,\cdot), p_{\varepsilon,\beta}(u;x,\cdot)) + F_{\varepsilon,\beta}(u;x)$$

is called in statistical mechanics the Gibbs free energy and we may write our energy as

$$\mathcal{E}_{\varepsilon,\beta}(u,w) = \sum_{x\in\widetilde{O}} G_{\varepsilon,\beta}(u,w;x).$$

With u fixed, the minimum of $\mathcal{E}_\varepsilon(u,w)$ with respect to w is attained when $w(x,\cdot) = p_{\varepsilon,\beta}(u;x,\cdot)$, i.e., $w(x,\cdot)$ coincides with the Boltzmann distribution for all x. Then

$$\mathcal{E}_{\varepsilon,\beta}(u,p_\beta) = \sum_{x\in\widetilde{O}} G_{\varepsilon,\beta}(u,p_{\varepsilon,\beta}(u);x) = \sum_{x\in\widetilde{O}} F_{\varepsilon,\beta}(u;x),$$

which is the sum of the free energies for all x. We will come back to this point of view later on in Section 6 when we look at the optimization algorithm in terms of the Expectation-Maximization (EM) algorithm.

4. Existence of minima. We shall consider here two cases: the patch NL-means and the patch NL-Poisson models. Since the existence of solutions for the patch NL-means model was already considered in [5] we only give a sketch of it here. The case of the patch NL-Poisson model will be considered in detail.

Let us introduce some notation. Let $C_c(\mathbb{R}^N)$ be the set of continuous functions with compact support in $\mathbb{R}^N$. By $C_c(\mathbb{R}^N)^+$ we denote the set of nonnegative functions in $C_c(\mathbb{R}^N)$. As usual, if Q is an open set we denote by $W^{1,p}(Q)$, $1 \le p \le \infty$, the

space of functions $v \in L^p(Q)$ such that $\nabla v \in L^p(Q)^N$. By $W^{1,p}(Q)^+$ we denote the set of nonnegative functions in $W^{1,p}(Q)$. We denote by $W^{2,p}(Q)$ (resp. by $W^{2,p}_{\rm loc}(Q)$), $1 \leq p \leq \infty$, the space of functions $v \in L^p(Q)$ such that $\nabla v \in L^p(Q)^N$ and $D^2 v \in L^p(Q)^{N\times N}$ (resp. the functions $v \in W^{2,p}(Q')$ for any subdomain Q' included in a compact set of Q).

Let us assume for the rest of the paper that $g \in L^1(\mathbb{R}^N)^+$ and $\int_{\mathbb{R}^N} g(h)\, dh = 1$.

4.1. Existence of minima for the patch NL-means model. We assume here that Ω is a rectangle in $\mathbb{R}^N$ and $\hat{u} : O^c \to \mathbb{R}$ with $\hat{u} \in L^\infty(O^c)$. We assume that $u : \Omega \to \mathbb{R}$ is such that $u|_{O^c} = \hat{u}$. We also assume that u is extended by symmetry and then by periodicity to $\mathbb{R}^N$.

In this Section we consider the patch NL-means model

$$\mathcal{E}_{2,T}(u,w) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x,y) \|p_u(x) - p_{\tilde{u}}(y)\|^2_{g,2} \mathrm{d}y\mathrm{d}x + T \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x,y) \log w(x,y) \mathrm{d}y\mathrm{d}x. \tag{4.1}$$

We implicitly understand that $\mathcal{E}_{2,T}(u,w) = +\infty$ in case that the second integral is not defined.

Let

$$\mathcal{W} := \{w \in L^1(\widetilde{O} \times \widetilde{O}^c) : \int_{\widetilde{O}^c} w(x,y)\, dy = 1 \text{ a.e. } x \in \widetilde{O}\}.$$

Let us consider the admissible class of functions

$$\mathcal{A}_2 := \{(u,w) : u \in L^\infty(\Omega),\, u = \hat{u} \text{ in } O^c,\, w \in \mathcal{W}\}.$$

Our purpose is to state the existence and regularity of minima of

$$\min_{(u,w)\in\mathcal{A}_2} \mathcal{E}_{2,T}(u,w) \tag{4.2}$$

and give a sketch of the proof. The existence was already proved in [5] and we refer to it for additional details.

PROPOSITION 4.1. *Assume that $g \in C_c(\mathbb{R}^N)^+$ has support contained in Ω_p, $\nabla g \in L^1(\mathbb{R}^N)$ and $\hat{u} \in BV(O^c) \cap L^\infty(O^c)$. Then there exists a minimum $(u,w) \in \mathcal{A}_2$ of $\mathcal{E}_{2,T}$. Moreover, for any minimum $(u,w) \in \mathcal{A}_2$ we have that $u \in W^{1,\infty}(O)$ and $w \in W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$.*

In other words, there are smooth minima and smooth probability distributions representing the fuzzy correspondences between $\widetilde{O}$ and $\widetilde{O}^c$. The proof of Proposition 4.1 is based on the following Lemmas whose proof can be found in [5].

LEMMA 4.2. *Assume that $g \in C_c(\mathbb{R}^N)^+$ has support contained in Ω_p, $\nabla g \in L^1(\mathbb{R}^N)$ and $\hat{u} \in BV(O^c) \cap L^\infty(O^c)$. Assume that $u \in L^\infty(\widetilde{O} + \Omega_p)$. Then the functions*

$$\nabla_x g * (u(x+\cdot) - \hat{u}(y+\cdot))^2 \quad \text{and} \quad \nabla_y g * (u(x+\cdot) - \hat{u}(y+\cdot))^2 \tag{4.3}$$

are uniformly bounded in $\widetilde{O} \times \widetilde{O}^c$ by a constant that depends on $\|\nabla g\|_{L^1}$, $\|u\|_\infty$, $\|\hat{u}\|_\infty$.

LEMMA 4.3. *Under the assumptions of Proposition 4.1, if $(u_n, w_n) \in \mathcal{A}_2$ is a minimizing sequence for $\mathcal{E}_{2,T}$ such that u_n is uniformly bounded, then we may extract a subsequence converging to a minimum of $\mathcal{E}_{2,T}$.*

This Lemma was proved in [5]. On the other hand, since a similar proof will be given in detail in next Section for the NL-Poisson model we omit the details here.

Proof of Proposition 4.1. The proof of existence follows from Lemma 4.3 once we observe that there are minimizing sequences $(u_n, w_n) \in \mathcal{A}_2$ for $\mathcal{E}_{2,T}$ such that u_n is uniformly bounded. For that, let $(u, w) \in \mathcal{A}_2$ and let us compute the equation satisfied by u if it is a minimum of $\mathcal{E}_{2,T}(\cdot, w)$. Indeed, since

$$\int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x,y) \|p_u(x) - p_{\hat{u}}(y)\|_{g,2}^2 \mathrm{d}y \mathrm{d}x = \int_{\mathbb{R}^N} \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z-\cdot, z'-\cdot))(u(z) - \hat{u}(z'))^2 \mathrm{d}z' \mathrm{d}z,$$

we have

$$u(z) = \int_{\mathbb{R}^N} g * (w_{\widetilde{O},\widetilde{O}^c}(z-\cdot, z'-\cdot))\hat{u}(z')\mathrm{d}z', \quad z \in O, \tag{4.4}$$

where we use the notation in (2.4), (2.5) and the fact that $k(w'; z) = 1$ for any $z \in O$ and any $w' \in \mathcal{W}$.

Similarly, if w is a minimum of $\mathcal{E}_{2,T}(u, \cdot)$ we have that

$$w(x,y) = w_{2,T}(u)(x,y) = \frac{1}{Z_{2,T}(u;x)} \exp\left(-\frac{1}{T} g * (u(x+\cdot) - \hat{u}(y+\cdot))^2\right). \tag{4.5}$$

Notice that both equations (4.4) and (4.5) hold if $(u, w) \in \mathcal{A}_2$ is a minimum of $\mathcal{E}_{2,T}$.

To prove the existence of minima of of $\mathcal{E}_{2,T}$, let $(u'_n, w'_n) \in \mathcal{A}_2$ be a minimizing sequence for this energy. Let

$$u_n = \arg\min_u \mathcal{E}_{2,T}(u, w'_n), \tag{4.6}$$

$$w_n = \arg\min_w \mathcal{E}_{2,T}(u_n, w). \tag{4.7}$$

Since

$$\mathcal{E}_{2,T}(u_n, w_n) \leq \mathcal{E}_{2,T}(u_n, w'_n) \leq \mathcal{E}_{2,T}(u'_n, w'_n),$$

$(u_n, w_n) \in \mathcal{A}_2$ is also a minimizing sequence of (4.2). By (4.4) we have

$$u_n(z) = \int_{\mathbb{R}^N} g * ((w'_n)_{\widetilde{O},\widetilde{O}^c}(z-\cdot, z'-\cdot))\hat{u}(z')\mathrm{d}z', \quad z \in O, \tag{4.8}$$

and we deduce that $\|u_n\|_{L^\infty(O)} \leq \|\hat{u}\|_\infty$. Then, using Lemma 4.3 we may extract a subsequence of (u_n, w_n) converging to a minimum of $\mathcal{E}_{2,T}$.

Let us prove the regularity assertion. Let $(u, w) \in \mathcal{A}_2$ be a minimum of $\mathcal{E}_{2,T}$. Let us first prove that $\nabla_x w(x,y)$ and $\nabla_y w(x,y)$ are bounded. By (4.4) we have that $\|u\|_{L^\infty(O)} \leq \|\hat{u}\|_\infty$ and

$$a \leq Z_{2,T}(u)(x) \leq b, \tag{4.9}$$

for some constants $b > a > 0$ which only depend on $\|\hat{u}\|_\infty$. Thus w is bounded and bounded away from zero. To abbreviate our expressions, let $U(x,y) := g * (u(x+\cdot) - \hat{u}(y+\cdot))^2$. Since

$$\nabla_x w(x,y) = -\frac{\exp\left(-\frac{1}{T}U(x,y)\right)}{Z_{2,T}(u)(x)} \left(\frac{1}{T}\nabla_x U(x,y) + \frac{\nabla_x Z_{2,T}(u)(x)}{Z_{2,T}(u)(x)}\right),$$

$$\nabla_y w(x,y) = -\frac{\exp\left(-\frac{1}{T}U(x,y)\right)}{T Z_{2,T}(u)(x)} \nabla_y U(x,y),$$

where

$$\nabla_x Z_{2,T}(u)(x) = -\frac{1}{T}\int_{\widetilde{O}^c} \exp\left(-\frac{1}{T}U(x,y)\right)\nabla_x U(x,y)dy,$$

by Lemma 4.2 and (4.9) we conclude that $\nabla_x w(x,y)$ and $\nabla_y w(x,y)$ are bounded.

Now, by (4.4), for any $x \in O$ we have

$$\nabla_x u(x) = \int_{\mathbb{R}^N} \nabla_x g * (w_{\widetilde{O},\widetilde{O}^c}(x-\cdot, z'-\cdot))\hat{u}(z')\mathrm{d}z'.$$

Since

$$\begin{aligned}\nabla_x g * (w_{\widetilde{O},\widetilde{O}^c}(x-\cdot, z'-\cdot)) = &\int_{\mathbb{R}^N} g(h)\nabla_x w(x+h, z'+h)\chi_{\widetilde{O}}(x+h)\chi_{\widetilde{O}^c}(z'+h)dh \\ &+ \int_{\mathbb{R}^N} g(h)w(x+h, z'+h)\nabla_x\chi_{\widetilde{O}}(x+h)\chi_{\widetilde{O}^c}(z'+h)dh\end{aligned}$$

and $\chi_{\widetilde{O}} \in BV(\mathbb{R}^N)$ we conclude that $\nabla_x g * (w_{\widetilde{O},\widetilde{O}^c}(x-\cdot, z'-\cdot))$ is bounded. We deduce that $\nabla_x u(x)$ is bounded. Hence $u \in W^{1,\infty}(O)$ and $w \in W^{1,\infty}(\widetilde{O}\times\widetilde{O}^c)$. ∎

4.2. Existence of minima for the patch NL-Poisson model. In this Section we consider the model

$$\mathcal{E}_{\nabla,T}(u,w) := \int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,\nabla}\mathrm{d}y\mathrm{d}x + T\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)\log w(x,y)\mathrm{d}y\mathrm{d}x, \tag{4.10}$$

where we denote $\|p\|^2_{g,\nabla} = \int_{\mathbb{R}^N} g(h)\|\nabla p(h)\|_2^2\mathrm{d}h$ for any $p \in W^{1,2}(\Omega_p)$. Recall that we assume that $u|_{O^c} = \hat{u}$.

Let

$$\mathcal{A}_\nabla := \{(u,w) \in \mathcal{A}_2 : u \in W^{1,2}(O),\ u|_{\partial O} = \hat{u}|_{\partial O^c}\}.$$

Our purpose is to prove the following result stating the existence of minima of

$$\min_{(u,w)\in\mathcal{A}_\nabla} \mathcal{E}_{\nabla,T}(u,w). \tag{4.11}$$

Proposition 4.4. *Assume that $\hat{u} \in W^{2,2}(O^c) \cap L^\infty(O^c)$ and $g \in W^{1,\infty}(\mathbb{R}^N)^+$ has compact support in Ω_p. There exists a solution of the variational problem (4.11). Moreover for any solution $(u,w) \in \mathcal{A}_\nabla$ we have $u \in W^{1,2}(O) \cap W^{2,p}_{loc}(O) \cap L^\infty(O)$ for all $p \in [1,\infty)$ and $w \in W^{1,\infty}(\widetilde{O}\times\widetilde{O}^c)$.*

Lemma 4.5. *Assume that $\hat{u} \in W^{1,2}(O^c)$ and $g \in L^\infty(\mathbb{R}^N)^+$ has compact support on Ω_p. Let $(u,w) \in \mathcal{A}_\nabla$. Assume that $\mathcal{E}_{\nabla,T}(u,w) \le C$. Then*

$$\|u\|_{W^{1,2}(O)} \le C'(C, \|\nabla\hat{u}\|_{L^2(O^c)}), \tag{4.12}$$

where $C' = C'(C, \|\nabla\hat{u}\|_{L^2(O^c)})$ denotes a constant that depends on its arguments.

Proof. Since the left hand side of

$$\begin{aligned}&\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)g * |\nabla_x u(x+\cdot) - \nabla_y\hat{u}(y+\cdot)|^2\,dxdy \\ &= \int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)\left(g * |\nabla_x u(x+\cdot)|^2 + \right. \\ &\qquad \left. + g * |\nabla_y\hat{u}(y+\cdot)|^2 - 2g * (\nabla_x u(x+\cdot)\cdot\nabla_y\hat{u}(y+\cdot))\right)dxdy\end{aligned}$$

is upper bounded,

$$|g * (\nabla_x u(x+\cdot)\cdot\nabla_y \hat{u}(y+\cdot))| \le \frac{\epsilon}{2} g * |\nabla_x u(x+\cdot)|^2 + \frac{1}{2\epsilon} g * |\nabla_y \hat{u}(y+\cdot)|^2,$$

for any $\epsilon > 0$, and

$$\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y) g * |\nabla_x \hat{u}(y+\cdot)|^2\, dxdy \le C'(C, \|\nabla \hat{u}\|_{L^2(O^c)}),$$

taking $\epsilon = \frac{1}{2}$, the result follows from

$$\begin{aligned}
&\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y) g * |\nabla_x u(x+\cdot)|^2\, dxdy \\
&= \int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y) \int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h)|^2\, dxdydh \\
&= \int_{\widetilde{O}}\int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h)|^2\, dxdh = \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \chi_{\widetilde{O}}(x) g(h)|\nabla_x u(x+h)|^2\, dx \\
&= \int_{\mathbb{R}^N}\int_{\mathbb{R}^N} \chi_{\widetilde{O}}(y-h) g(h)|\nabla_x u(y)|^2\, dhdy \ge \int_O\int_{\mathbb{R}^N} \chi_{\widetilde{O}}(y-h) g(h)|\nabla_x u(y)|^2\, dhdy \\
&\ge \int_O\int_{\mathbb{R}^N} g(h)|\nabla_x u(y)|^2\, dhdy = \int_O |\nabla_x u(y)|^2\, dy.
\end{aligned}$$

□

LEMMA 4.6. *Assume that $\hat{u} \in W^{2,2}(O^c)$, $u \in W^{1,2}(O)$, $u|_{\partial O} = \hat{u}|_{\partial O^c}$ and $g \in W^{1,\infty}(\mathbb{R}^N)^+$ has compact support on Ω_p. Then*

$$\begin{aligned}
&\nabla_x \int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h) - \nabla_y \hat{u}(y+h)|^2\, dh \quad \text{and} \\
&\nabla_y \int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h) - \nabla_y \hat{u}(y+h)|^2\, dh
\end{aligned} \tag{4.13}$$

are bounded in $L^\infty(\widetilde{O}\times\widetilde{O}^c)$ with a bound depending on $\|\hat{u}\|_{W^{2,2}(O^c)}$, $\|g\|_{W^{1,\infty}}$ and $\|\nabla u\|_{L^2(O)}$.

Proof. The bounds in (4.13) follow if we prove that

$$\nabla_{x+y} \int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h) - \nabla_y \hat{u}(y+h)|)^2\, dh \quad \text{and}$$

$$\nabla_{x-y} \int_{\mathbb{R}^N} g(h)|\nabla_x u(x+h) - \nabla_y \hat{u}(y+h)|^2\, dh$$

are bounded in L^∞ with a bound depending on $\|\hat{u}\|_{W^{2,2}(O^c)}$, $\|g\|_{W^{1,\infty}}$ and $\|\nabla u\|_{L^2(\tilde{O})}$. Let us write $u_i(x) = \partial_{x_i} u(x)$, $\hat{u}_i(x) = \partial_{x_i}\hat{u}(x)$, $i = 1,\dots,N$. Let $x \in \widetilde{O}$, $y \in \widetilde{O}^c$, $i \in \{1,\dots,N\}$. As in the proof of Lemma 4.2 we have

$$\begin{aligned}
&\nabla_{x+y} \int_{\mathbb{R}^N} g(h)(u_i(x+h) - \hat{u}_i(y+h))^2\, dh \\
&= 2\int_{\mathbb{R}^N} g(h)(u_i(x+h) - \hat{u}_i(y+h))(\nabla_x u_i(x+h) - \nabla_y \hat{u}_i(y+h))\, dh \\
&= 2\int_{\mathbb{R}^N} g(h)(u_i(x+h) - \hat{u}_i(y+h))(\nabla_h u_i(x+h) - \nabla_h \hat{u}_i(y+h))^2\, dh \\
&= -\int_{\mathbb{R}^N} \nabla_h g(h)(u_i(x+h) - \hat{u}_i(y+h))^2\, dh.
\end{aligned}$$

Then

$$\left|\nabla_{x+y}\int_{\mathbb{R}^N} g(h)(u_i(x+h)-\hat{u}_i(y+h))^2\,dh\right| \\ \leq 2\|\nabla_h g\|_\infty\left(\int_{\tilde{O}+\Omega_p}|\nabla u(y)|^2dy+\int_{\widetilde{O}^c+\Omega_p}|\nabla\hat{u}(y)|^2dy\right).$$

Notice that since $u\in W^{1,2}(O)$, $u|_{\partial O}=\hat{u}|_{\partial O^c}$ and $u=\hat{u}$ on O^c we have that $u\in W^{1,2}(\tilde{O}+\Omega_p)$.

As above, by direct computation and after integration by parts, we have

$$\begin{aligned}&\nabla_{x-y}\int_{\mathbb{R}^N} g(h)(u_i(x+h)-\hat{u}_i(y+h))^2\,dh\\ &=-\int_{\mathbb{R}^N}\nabla_h g(h)(u_i(x+h)^2-\hat{u}_i(y+h)^2)\,dh\\ &+2\int_{\mathbb{R}^N} g(h)u_i(x+h)\nabla_h\hat{u}_i(y+h)\,dh+2\int_{\mathbb{R}^N}\nabla_h(g(h)\hat{u}_i(y+h))u_i(x+h)\,dh.\end{aligned}$$

The three terms are bounded by a constant depending on $\|\hat{u}\|_{W^{2,2}(O^c)}$, $\|g\|_{W^{1,\infty}}$ and $\|\nabla u\|_{L^2(\tilde{O})}$. □

Proof of Proposition 4.4. Existence. Let (u_n,w_n) be a minimizing sequence of (4.11). Since Ω is a bounded domain we have that

$$\int_{\widetilde{O}}\int_{\widetilde{O}^c}\chi_{\{w_n>1\}}w_n(x,y)\log w_n(x,y)\mathrm{d}y\mathrm{d}x$$

is bounded. Hence $w_n(1+\log^+ w_n)$ is bounded in $L^1(\widetilde{O}\times\widetilde{O}^c)$. Then the sequence w_n is relatively weakly compact in L^1 and modulo a subsequence we may assume that w_n weakly converges in $L^1(\widetilde{O}\times\widetilde{O}^c)$ to some $w\in\mathcal{W}$.

By Lemma 4.5, we have that u_n is uniformly bounded in $W^{1,2}(\widetilde{O})$. By Lemma 4.6, we have that

$$\nabla_x\int_{\mathbb{R}^N} g(h)|\nabla u(x+h)-\nabla u(y+h)|^2\,dh\quad\text{and}\quad\nabla_y\int_{\mathbb{R}^N} g(h)|\nabla u(x+h)-\nabla u(y+h)|^2\,dh \tag{4.14}$$

are uniformly bounded in $L^\infty(\widetilde{O}\times\widetilde{O}^c)$. Thus, modulo the extraction of a subsequence, we may assume that $u_n\to u$ a.e. and in $L^2(\widetilde{O})$, $\nabla u_n\to\nabla u$ weakly in $L^2(\widetilde{O}+\Omega_p)$, and $g*(\nabla_x u_n(x+\cdot)-\nabla_y\hat{u}_n(y+\cdot))^2$ converges strongly in all L^p spaces, $1\leq p<\infty$, and also in the dual of $L\,Log^+L$ to some function W. Then by passing to the limit as $n\to\infty$ we have

$$\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)W(x,y)\mathrm{d}y\mathrm{d}x+T\int_{\widetilde{O}}\int_{\widetilde{O}^c} w(x,y)\log w(x,y)\mathrm{d}y\mathrm{d}x\leq\liminf_n\mathcal{E}_{\nabla,T}(u_n,w_n).$$

Taking test functions $\psi(x,y)$, integrating in $\widetilde{O}\times\widetilde{O}^c$ and using the convexity of the square function, we have

$$\int g(h)(\nabla_x u(x+h)-\nabla_y\hat{u}(y+h))^2\,dh\leq W(x,y).$$

Thus

$$\mathcal{E}_{\nabla,T}(u,w)\leq\liminf_n\mathcal{E}_{\nabla,T}(u_n,w_n).$$

Regularity. Observe that if $(u, w) \in \mathcal{A}_\nabla$ is a minimum of (4.11), then $u \in W^{1,2}(O)$ (by Lemma 4.5) and it satisfies the Euler-Lagrange equations. By fixing w and computing the first variation of $\mathcal{E}_{\nabla,T}$ with respect to its first variable, we have that u is a solution of the Poisson equation (2.6). On the other hand, if u is fixed and we compute the first variation of $\mathcal{E}_{\nabla,T}$ with respect to its second variable, then

$$w(x,y) = w_{\nabla,T}(u;x,y) = \frac{1}{Z_{\nabla,T}(u;x)} \exp\left(-\frac{1}{T}\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,\nabla}\right), \tag{4.15}$$

where the normalizing factor $Z_{\nabla,T}(u)$ is given by

$$Z_{\nabla,T}(u;x) = \int_{\widetilde{O}^c} \exp\left(-\frac{1}{T}\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,\nabla}\right) \mathrm{d}y. \tag{4.16}$$

As in the proof of Proposition 4.1 we observe that $Z_{\nabla,T}(u;x)$ is bounded and bounded away from zero (thanks to Lemmas 4.5 and 4.6). Then, it follows that $w \in W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$. Using this and (2.7) we have that $\boldsymbol{v}(w) \in W^{1,\infty}(O)^2$. Then the solution u of (2.6) is in $W^{1,2}(O) \cap W^{2,p}_{\mathrm{loc}}(O) \cap L^\infty(O)$ for any $p \in [1,\infty)$ [29]. ∎

Remark 1. Notice that the regularity result holds for any $(u^*, w^*) \in \mathcal{A}_\nabla$ satisfying (2.6) and (4.15).

5. Optimal correspondence maps. In this Section we consider a relaxation of the correspondence map approach that appears as the Gamma limit of (1.2). In turn, the energy (1.2) can be considered as a relaxation of the energy considered in [22] (see also [6]). We prove existence of solutions of the relaxed problem and the existence of optimal correspondence maps. Then in Subsection 5.2 we prove the existence of correspondence maps which are uniform limits of bounded variation functions with finitely many values. We give the details corresponding to the patch NL-means model. Analogous results with similar proofs hold for the patch NL-Poisson model (see Remark 3).

5.1. Existence of optimal correspondence maps. Let us first recall the notion of measurable measure-valued map.

Definition 5.1 (Measurable measure-valued map). *Let $\mathcal{X} \subseteq \mathbb{R}^N$, $\mathcal{Y} \subseteq \mathbb{R}^M$ be open sets, μ be a positive Radon measure in $\mathcal{X}$ and $x \to \nu_x$ be a function that assigns to each x in $\mathcal{X}$ a Radon measure ν_x on $\mathcal{Y}$. We say that the map is μ-measurable if $x \to \nu_x(B)$ is μ-measurable for any Borel set B in $\mathcal{Y}$.*

By the disintegration theorem, if ν is a positive Radon measure in $\mathcal{X} \times \mathcal{Y}$ such that $\nu(K \times \mathcal{Y}) < \infty$ for any compact set $K \subseteq \mathcal{X}$ and $\mu = \pi_\sharp \nu$ where $\pi : \mathcal{X} \times \mathcal{Y} \to \mathcal{X}$ is the projection on the first factor (*i.e.* $\mu(B) = \nu(B \times Y)$ for any Borel set $B \subseteq X$), then there exist a measurable measure-valued map $x \to \nu_x$ such that $\nu_x(\mathcal{Y}) = 1$ μ-a.e. in $\mathcal{X}$ and for any $\psi \in L^1(\mathcal{X} \times \mathcal{Y}, \nu)$ we have

$$\begin{aligned}
&\psi(x,\cdot) \in L^1(\mathcal{Y}, \nu_x) \qquad \text{for } \mu\text{-a.e. } x \in \mathcal{X},\\
&x \to \int_{\mathcal{Y}} \psi(x,y)\, d\nu_x(y) \in L^1(\mathcal{X}, \mu),\\
&\int_{\mathcal{X}\times\mathcal{Y}} \psi(x,y)\, d\nu(x.y) = \int_{\mathcal{X}} \int_{\mathcal{Y}} \psi(x,y)\, d\nu_x(y)\, d\mu(x).
\end{aligned}$$

Let us consider $\mathcal{MP}$ the set of measurable measure valued maps $\nu \geq 0$ in $\widetilde{O} \times \mathrm{cl}(\widetilde{O}^c)$ such that $\pi_\sharp \nu = \mathcal{L}^N|_{\widetilde{O}}$, where $\mathcal{L}^N|_{\widetilde{O}}$ denotes the Lebesgue measure restricted

to $\widetilde{O}$. We assume that $g \in C_c(\mathbb{R}^N)$ has support contained in Ω_p, $\nabla g \in L^1(\mathbb{R}^N)$ and $\hat{u} \in BV(O^c) \cap L^\infty(O^c)$. Let

$$\mathcal{A}_{2,0} := \{(u,\nu) : u \in L^\infty(\Omega),\ u = \hat{u} \text{ in } O^c,\ \nu \in \mathcal{MP}\}.$$

For $(u,\nu) \in \mathcal{A}_{2,0}$, define

$$\mathcal{E}_{2,0}(u,\nu) := \int_{\widetilde{O}} \int_{\widetilde{O}^c} g * (u(x+\cdot) - \hat{u}(y+\cdot))^2 d\nu(x,y). \tag{5.1}$$

Notice that, by Lemma 4.2, the above integral is well defined. We are now ready to state the existence result.

PROPOSITION 5.2. *There exists a minimum $(u,\nu) \in \mathcal{A}_{2,0}$ of $\mathcal{E}_{2,0}$.*

Proof. Let $(u_n,\nu_n) \in \mathcal{A}_{2,0}$ be a minimizing sequence of $\mathcal{E}_{2,0}$. Let us observe that we may take u_n uniformly bounded. Indeed, let $\kappa = 2\|\hat{u}\|_\infty$ and $u_{\kappa,n} = u_n \chi_{|u_n| \le \kappa}$. Then

$$|u_{\kappa,n}(x) - \hat{u}(y)| \le |u_n(x) - \hat{u}(y)|.$$

This is clearly true if $|u_n(x)| \le \kappa$. Now if $|u_n(x)| > \kappa$ we have

$$|u_n(x) - \hat{u}(y)| > \kappa - \hat{u}(y) \ge \|\hat{u}\|_\infty \ge |u_{\kappa,n}(x) - \hat{u}(y)|.$$

Thus $g * (u_{\kappa,n}(x+\cdot) - \hat{u}(y+\cdot))^2 \le g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2$ and $(u_{\kappa,n},\nu_n)$ is also a minimizing sequence. Thus, we may assume that u_n is uniformly bounded. By extracting a subsequence, if necessary, we may assume that u_n converges to some $u \in L^\infty(\widetilde{O} + \Omega_p)$ with $u|_{O^c} = \hat{u}$. By Lemma 4.2, $g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2$ is uniformly Lipschitz in x, y and we may assume that it converges uniformly to a function $W(x,y)$. On the other hand, we may also assume that $\nu_n \to \nu$ weakly* as measures. Hence

$$\int_{\widetilde{O}} \int_{\widetilde{O}^c} W(x,y) d\nu(x,y) \le \liminf_n \int_{\widetilde{O}} \int_{\widetilde{O}^c} g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2 d\nu(x,y).$$

Notice that $\pi_\sharp \nu = \mathcal{L}^N|_{\widetilde{O}}$. Clearly from the weak* convergence, we have that $\nu(V \times \mathrm{cl}(\widetilde{O}^c)) \le \mathcal{L}^N(V)$ for any open set $V \subseteq \widetilde{O}$ and $\nu(K \times \mathrm{cl}(\widetilde{O}^c)) \ge \mathcal{L}^N(K)$ for any compact set $K \subseteq \widetilde{O}$. By regularity of the measure ν we have that $\nu(B \times \mathrm{cl}(\widetilde{O}^c)) = |B|$ for any Borel set $B \subseteq \widetilde{O}$. Thus $\pi_\sharp \nu = \mathcal{L}^N|_{\widetilde{O}}$.

Now, as in the proof of Proposition 4.1, we have that $g * (u(x+\cdot) - \hat{u}(y+\cdot))^2 \le W(x,y)$. Hence

$$\int_{\widetilde{O}} \int_{\widetilde{O}^c} g*(u(x+\cdot)-\hat{u}(y+\cdot))^2 d\nu(x,y) \le \liminf_n \int_{\widetilde{O}} \int_{\widetilde{O}^c} g*(u_n(x+\cdot)-\hat{u}(y+\cdot))^2 d\nu(x,y).$$

Thus $(u,\nu) \in \mathcal{A}_{2,0}$ is a minimum of $\mathcal{E}_{2,0}$. □

REMARK 2. If $\hat{u} \ge 0$ we may also take $u_{\kappa,n} = \tau_\kappa(u_n)$ where $\kappa \ge \|\hat{u}\|_\infty$ and $\tau_\kappa(r) = \mathrm{sign}(r) \inf(|r|,\kappa)$.

Let $\varphi : \widetilde{O} \to \widetilde{O}^c$ be a measurable map. Then $x \in \widetilde{O} \to \nu_x = \delta_{\varphi(x)}(y)$ is measurable. Similarly if the map $x \in \widetilde{O} \to \nu_x = \delta_{\varphi(x)}(y)$ is measurable then φ is measurable. Let us denote by ν^φ the measure determined by φ.

PROPOSITION 5.3. *There exists a minimum $(u^*,\nu^*) \in \mathcal{A}_{2,0}$ of $\mathcal{E}_{2,0}$ such that $\nu^* = \nu^\varphi$ where $\varphi : \widetilde{O} \to \mathrm{cl}(\widetilde{O}^c)$ is a measurable map.*

Proof. Let $(u^*, \nu^*) \in \mathcal{A}_{2,0}$ be a minimum of $\mathcal{E}_{2,0}$. Then $\nu^* \in \arg\min_{\nu \in \mathcal{MP}} \mathcal{E}_{2,0}(u^*, \nu)$. Let us prove that there is a measurable map φ such that ν^φ is a minimum of $\mathcal{A}_{2,0}$. Indeed, this is a consequence of the Kuratowski-Ryll-Nardzewski Theorem [52] (Theorem 5.2.1) or [2] (Theorem 14.86). Let us consider $U(x,y) = g * (u(x+\cdot) - \hat{u}(y+\cdot))^2$, $x \in \widetilde{O}$, $y \in \widetilde{O}^c$. Let

$$m(x) = \min_{y \in \widetilde{O}^c} U(x,y),$$
$$M(x) = \{y \in \widetilde{O}^c : U(x,y) = m(x)\}.$$

Then, by Berge's maximum theorem applied to $-U(x,y)$ [2] (Theorem 14.30) we have that $m(x)$ is continuous and $M(x)$ is an upper hemicontinuous correspondence with compact values. Thus it has a closed graph [2] (Theorem 14.12), hence it is measurable [2] (Theorem 14.68). By the Kuratowski-Ryll-Nardzewski Theorem [52] (Theorem 5.2.1) or [2] (Theorem 14.86), we know that $x \to M(x)$ admits a measurable selector, that is there is a measurable map $x \to \varphi(x) \in M(x)$. The measure ν^φ is a minimum of $\mathcal{A}_{2,0}$ since we may write

$$U(x, \varphi(x)) = \min_{y \in \widetilde{O}^c} U(x,y) \leq \int_{\widetilde{O}^c} U(x,y) d\nu_x^*(y).$$

Integrating in x we deduce that

$$\mathcal{E}_{2,0}(u^*, \nu^\varphi) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} U(x,y) d\nu_x^\varphi(y) \mathrm{d}x \leq \int_{\widetilde{O}} \int_{\widetilde{O}^c} U(x,y) d\nu_x^*(y) \mathrm{d}x = \mathcal{E}_{2,0}(u^*, \nu^*).$$

□

Let us sketch a second proof which gives a different point of view. Clearly, since the energy function $\nu \to \mathcal{E}_{2,0}(u^*, \nu)$ is linear in ν, there are minima that are attained on the set of extreme points of the convex set $\mathcal{MP}$. Thus, the proposition is a consequence of the following Lemma whose proof will be given in the Appendix.

LEMMA 5.4. *The set of extreme points of the convex set $\mathcal{MP}$ coincides with the set of measures $\{\nu^\varphi : \varphi$ is a measurable map$\}$.*

We now address the relation between the patch NL-means functional for $T > 0$ and $\mathcal{E}_{2,0}$.

PROPOSITION 5.5. *The energies $\mathcal{E}_{2,T}$ Gamma-converge to the energy $\mathcal{E}_{2,0}$. In particular, the minima of $\mathcal{E}_{2,T}$ converge to minima of $\mathcal{E}_{2,0}$.*

Proof. Let $(u, \nu) \in \mathcal{A}_{2,0}$ and $(u_n, w_n) \in \mathcal{A}_2$ be such that $u_n \to u$ weakly* in L^∞ and $w_n \to \nu$ weakly* as measures. The fact that $u_n \to u$ weakly* in L^∞ includes that u_n is uniformly bounded. By Lemma 4.2, $g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2$ is uniformly Lipschitz and converges uniformly to some function $W(x,y)$. Then

$$\int_{\widetilde{O}} \int_{\widetilde{O}^c} W(x,y) d\nu(x,y) \leq \liminf_n \int_{\widetilde{O}} \int_{\widetilde{O}^c} w_n(x,y) g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2 dxdy$$

$$\leq \liminf_n \int_{\widetilde{O}} \int_{\widetilde{O}^c} w_n(x,y) g * (u_n(x+\cdot) - \hat{u}(y+\cdot))^2 dxdy + T_n \int_{\widetilde{O}} \mathcal{H}(w_n(x,\cdot))\, dx.$$

Since $g * (u(x+\cdot) - \hat{u}(y+\cdot))^2 \leq W(x,y)$, we have that

$$\mathcal{E}_{2,0}(u, \nu) \leq \liminf_n \mathcal{E}_{2,T_n}(u_n, w_n).$$

Now, let $(u,\nu)\in\mathcal{A}_{2,0}$. Let ν_x be the probability measures obtained by disintegrating ν with respect to $\mathcal{L}^N|_{\widetilde{O}}$. Let us take $w_n(x,y)=g_{T_n}*\nu_x(y)$ where $g_T(y)=\frac{1}{T^2}g(\frac{y}{T})$ for any $T>0$. Thus $w_n\in\mathcal{W}$, $w_n\to\nu$ weakly* as measures and $w_n\leq\frac{1}{T_n^N}\|g\|_\infty$. This implies that $(u,w_n)\in\mathcal{A}_2$ and

$$\lim_n T_n\int_{\widetilde{O}}\int_{\widetilde{O}^c} w_n(x,y)\log\, w_n(x,y)\,dxdy=0.$$

On the other hand

$$\lim_n\int_{\widetilde{O}}\int_{\widetilde{O}^c} w_n(x,y)g*(u(x+\cdot)-\hat{u}(y+\cdot))^2\,dxdy=\int_{\widetilde{O}}\int_{\widetilde{O}^c} g*(u(x+\cdot)-\hat{u}(y+\cdot))^2\,d\nu(x,y).$$

Thus $\limsup_n\mathcal{E}_{2,T_n}(u,w_n)\leq\mathcal{E}_{2,0}(u,\nu)$. □

Remark 3. Statements analogous to Propositions 5.2, 5.3 and 5.5 also hold for the non-local Poisson model where the limit energy is now

$$\mathcal{E}_{\nabla,0}(u,\nu):=\int_{\widetilde{O}}\int_{\widetilde{O}^c} g*(\nabla u(x+\cdot)-\nabla\hat{u}(y+\cdot))^2 d\nu(x,y).$$

In this case, we assume that $\hat{u}\in W^{2,2}(O^c)$ and $g\in W^{1,\infty}(\mathbb{R}^N)$ has compact support in Ω_p and we use Lemmas 4.5 and 4.6.

5.2. Regularity of optimal correspondence maps. The following result gives us some information on the the existence of correspondences with some regularity.

Theorem 5.6. *Let X be an open bounded subset of $\mathbb{R}^N$ with Lipschitz boundary and Y be a compact subset of $\mathbb{R}^m$. Let $U:X\times Y\to\mathbb{R}$ be a Lipschitz continuous function. For each $x\in X$, let $M(x):=\{y\in Y: U(x,y)=\min_{\bar{y}\in Y}U(x,\bar{y})\}$. Then there exists a selection of the multifunction $x\in X\to M(x)\subseteq Y$, i.e., a function $S:X\to Y$ such that $S(x)\in M(x)\ \forall x\in X$, which is a uniform limit of functions in $BV(X)^m$.*

The result can be immediately applied to our case with $X=\widetilde{O}$, $Y=\widetilde{O}^c$. Indeed, it can be applied to the patch NL-means model when we are under the assumptions of Lemma 4.2 and to the patch non-local Poisson model when the assumptions of Proposition 4.4 hold. Theorem 5.6 implies that the offsets map $\tau(x)=\varphi(x)-x$ is a uniform limit of maps of bounded variation (in this case from $\widetilde{O}$ to $\mathbb{R}^2$), since the identity map $x\to x$ is also.

Notice that the result does not say that all optimal correspondence maps are regular. In view of Propositions 4.1 and 4.4, this raises the question if the solution obtained by annealing, i.e. by solving $\mathcal{E}_{2,T}$ (or $\mathcal{E}_{\nabla,T}$) and letting $T\to 0+$ is indeed a regular solution in the sense described in Theorem 5.6. We are not able to answer this question, at present.

We included a brief summary of the properties of BV functions after the proof of this theorem. Notice that we defined X as an open set since we define the BV space on open sets. The same statement holds true if we replace X by its closure $\overline{X}$. In that case, since X has a Lipschitz boundary (hence of Lebesgue null measure) functions in $BV(X)$ uniquely determine its extension to $\overline{X}$ (the trace on ∂X is well defined) [3].

Its proof is a simple adaptation of the Kuratowski-Ryll-Nardzewski Theorem [52] (Theorem 5.2.1) or [2] (Theorem 14.86). For that we need the following simple Lemma.

Lemma 5.7. *Let $m(x)=\inf_{y\in Y}U(x,y)$, $x\in X$. The function $m(x)$ is Lipschitz. Also are the functions $x\in X\to\inf_{\bar{y}\in B}U(x,\bar{y})$ for any $B\subseteq Y$ and $\epsilon>0$.*

Proof. Let $x, \bar{x} \in X$. Since U is Lipschitz in both variables we have

$$m(x) \leq U(x,y) \leq U(\bar{x},y) \leq L|x-\bar{x}|.$$

Taking infima with respect to y in the right hand side we have $m(x) \leq m(\bar{x}) + L|x-\bar{x}|$. By symmetry we have that m is Lipschitz. The more general case follows in the same way. □

Proof of Theorem 5.6. Let

$$M_\epsilon(x) = \{y \in Y : U(x,y) \leq \min_{\bar{y}\in Y} U(x,\bar{y}) + \epsilon\}, \quad x \in X.$$

Let $\epsilon_n \downarrow 0$ and $M_n(x) := M_{\epsilon_n}(x)$. Let d denote the euclidian distance in Y. Without loss of generality, assume that $d < 1$.

We define inductively a sequence $S_n : X \to Y$ of functions in $BV(X)^m$ taking finitely many values such that for every $x \in X$ and every $n \in \mathbb{N}$,

(i) $d(S_n(x), M(x)) \leq 2^{-n}$, and
(ii) $d(S_{n-1}(x), S_n(x)) \leq 2^{-n+2}$.

For that, let $\{r_k\}$ be a countable dense set in Y. Define $S_0(x) = r_0$ for all $x \in X$. Let $n > 0$ and assume that we have constructed S_i satisfying (i), (ii) for all $i < n$. For each $k \in \mathbb{N}$, let

$$E_k^n = \{x \in X : |S_{n-1}(x) - r_k| \leq 2^{-n+2},\ d(r_k, M_n(x)) \leq 2^{-n}\}.$$

Let us prove that for a suitable selection of ϵ_n the sets E_k^n are sets of finite perimeter in X. Since S_{n-1} is in $BV(X)^m$ and takes finitely many values, the set $\{x \in X : |S_{n-1}(x) - r_k| \leq 2^{-n+2}\}$ is of finite perimeter in X. Observe

$$\begin{aligned}\{x \in X : d(r_k, M_n(x)) \leq 2^{-n}\} &= \{x \in X : \exists y \in B(r_k, 2^{-n}) \text{ s. t. } U(x,y) \leq m(x) + \epsilon_n\} \\ &= \{x \in X : \inf_{y \in B(r_k, 2^{-n})} U(x,y) - m(x) \leq \epsilon_n\}.\end{aligned}$$

Since by Lemma 5.7 the functions $\inf_{y\in B(r_k,2^{-n})} U(x,y)$ and $m(x)$ are Lipschitz functions of x, by excluding a set of null measure we may select a sequence $\epsilon_n \to 0+$ such that all sets E_k^n are sets of finite perimeter in X.

Let us consider a finite set $R_n \subset \{r_k\}_k$, which is a 2^{-n} net in Y, that is, any point $y \in Y$ is at distance at most 2^{-n} from R_n. Let us prove that $X = \cup_{k:r_k\in R_n} E_k^n$. Let $x \in X$. Since, by (i), $d(S_{n-1}(x), M(x)) \leq 2^{-n+1}$, there exists $y \in M(x)$ such that $|S_{n-1}(x) - y| \leq 2^{-n+1}$. Let $r_k \in R_n$ be such that

$$|y - r_k| \leq 2^{-n}. \tag{5.2}$$

Then

$$|S_{n-1}(x) - r_k| \leq |S_{n-1}(x) - y| + |y - r_k| \leq 2^{-n+1} + 2^{-n} < 2^{-n+2}. \tag{5.3}$$

Thus $x \in E_k^n$.

Then there exist pairwise disjoint sets $D_k^n \subseteq E_k^n$ such that $\cup_{k:r_k\in R_n} D_k^n = \cup_{k:r_k\in R_n} E_k^n = X$ and D_k^n are sets of finite perimeter in X. It suffices to take $D_k^n = E_k^n \setminus \cup_{j<k} E_j^n$.

Let $S_n(x) = \sum_{k:r_k\in R_n} r_k \chi_{D_k^n}(x)$. Clearly $S_n \in BV(X)^m$ and takes finitely many values in Y. By (5.2), we have that $d(S_n(x), M(x)) \leq 2^{-n}$. By (5.3), we have $d(S_{n-1}(x), S_n(x)) \leq 2^{-n+2}$.

Now, by (ii), the sequence S_n converges uniformly to some function $S : X \to Y$. By (i), $S(x) \in M(x)$ for all $x \in X$. Thus S is a uniform limit of functions in $BV(X)^m$. □

REMARK 4. Notice that the above proof also shows that given $\epsilon > 0$ there is a function $S_\epsilon \in BV(X)^m$ with finitely many values such that $U(x, S_\epsilon(x)) \leq m(x) + \epsilon$.

Let us point out some of the properties of S as a uniform limit of $BV(X)^m$ functions. For a detailed account of the properties of functions of bounded variation, we refer to [3].

Let Q be an open subset of $\mathbb{R}^N$. Let $u \in L^1_{\text{loc}}(Q)$. The total variation of u in Q is defined by

$$V(u, Q) := \sup \left\{ \int_Q u \ \text{div}\, \sigma \ dx : \sigma \in C_c^\infty(Q; \mathbb{R}^N), |\sigma(x)| \leq 1 \ \forall x \in Q \right\}, \tag{5.4}$$

where $C_c^\infty(Q; \mathbb{R}^N)$ denotes the vector fields with values in $\mathbb{R}^N$ which are infinitely differentiable and have compact support in Q. For a vector $v = (v_1, \ldots, v_N) \in \mathbb{R}^N$ we denoted $|v|^2 := \sum_{i=1}^N v_i^2$. Following the usual notation, we will denote $V(u, Q)$ by $|Du|(Q)$.

Let $u \in L^1(Q)$. We say that u is a *function of bounded variation* in Q if $V(u, Q) < \infty$. The vector space of functions of bounded variation in Q will be denoted by $BV(Q)$. Recall that $BV(Q)$ is a Banach space when endowed with the norm $\|u\| := \int_Q |u| \ dx + |Du|(Q)$.

A measurable set $E \subseteq Q$ is said to be of *finite perimeter* in Q if $\chi_E \in BV(Q)$. The perimeter of E in Q is defined as $P(E, Q) := |D\chi_E|(Q)$. Recall that almost all level sets of a bounded variation function are sets of finite perimeter.

Let us denote by $\mathcal{L}^N$ and $\mathcal{H}^{N-1}$, respectively, the N-dimensional Lebesgue measure and the $(N-1)$-dimensional Hausdorff measure in $\mathbb{R}^N$.

Let $u \in [L^1_{\text{loc}}(Q)]^m$ $(m \geq 1)$. We say that u has an approximate limit at $x \in Q$ if there exists $\xi \in \mathbb{R}^m$ such that

$$\lim_{\rho \downarrow 0} \frac{1}{|B(x, \rho)|} \int_{B(x,\rho)} |u(y) - \xi| dy = 0. \tag{5.5}$$

The set of points where this does not hold is called the approximate discontinuity set of u, and is denoted by S_u. Using Lebesgue's differentiation theorem, one can show that the approximate limit ξ exists at $\mathcal{L}^N$-a.e. $x \in Q$, and is equal to $u(x)$: in particular, $|S_u| = 0$. If $x \in Q \setminus S_u$, the vector ξ is uniquely determined by (5.5) and we denote it by $\tilde{u}(x)$.

We say that $x \in Q$ is an *approximate jump point of* u if there exist $u^+(x) \neq u^-(x) \in \mathbb{R}^m$ and $|\nu_u(x)| = 1$ such that

$$\lim_{\rho \downarrow 0} \frac{1}{|B^\pm_\rho(x, \nu_u(x))|} \int_{B^\pm_\rho(x, \nu_u(x))} |u(y) - u^\pm(x)| \, dy = 0$$

where $B^\pm_\rho, \nu_u(x)) = \{y \in B(x, \rho) \ : \ \pm\langle y - x, \nu_u(x)\rangle > 0\}$. We denote by J_u the set of approximate jump points of u. If $u \in BV(Q)^m$, the set S_u is countably $\mathcal{H}^{N-1}$ rectifiable, J_u is a Borel subset of S_u and $\mathcal{H}^{N-1}(S_u \setminus J_u) = 0$ [3]. In particular, we have that $\mathcal{H}^{N-1}$-a.e. $x \in Q$ is either a point of approximate continuity of $\tilde{u}$, or a jump point with two limits in the above sense.

As a uniform limit of functions in $BV(X)^m$, S inherits some of its properties. In particular, $\mathcal{H}^{N-1}$-a.e. $x \in X$ is either a point of approximate continuity of S, or a jump point with two limits in the above sense. Moreover given a point $x \in J_S$, $x \in J_{S_n}$ for n large enough. That is $J_S = \cap_k \cup_{n\geq k} J_{S_n} = \limsup_n J_{S_n}$. In particular, J_S is a countably rectifiable set [3]. Outside it, the function is approximately continuous modulo an $\mathcal{H}^{N-1}$ null set.

6. Convergence of the alternating optimization scheme. To minimize the energy, we use an alternating minimization scheme. At each iteration, two optimization steps are solved: the minimization of $\mathcal{E}$ with respect to w while keeping u fixed; and the minimization with respect to u with w fixed. In this Section we prove the convergence of such a scheme to a critical point of the energy both for the case of patch NL-means and -Poisson models.

Let $\mathcal{E}_{\varepsilon,T}$ be one of the energies $\mathcal{E}_{2,T}$ or $\mathcal{E}_{\nabla,T}$. Similarly, $\mathcal{A}_\varepsilon$ denotes $\mathcal{A}_2$ or $\mathcal{A}_\nabla$

Algorithm 1 Alternate minimization of $\mathcal{E}_{\varepsilon,T}$.

Initialization: choose u^0 with $\|u^0\|_\infty \leq \|\hat{u}\|_\infty$.

For each $k \in \mathbb{N}$ solve

$$w^{k+1} = \operatorname*{arg\,min}_{w\in\mathcal{W}} \mathcal{E}_{\varepsilon,T}(u^k, w), \tag{6.1}$$

$$u^{k+1} = \operatorname*{arg\,min}_{(u,w^{k+1})\in\mathcal{A}_\varepsilon} \mathcal{E}_{\varepsilon,T}(u, w^{k+1}), \tag{6.2}$$

Proposition 6.1. *The iterated optimization algorithm converges (modulo a subsequence) to a critical point $(u^*, w^*) \in \mathcal{A}_\varepsilon$ of $\mathcal{E}_{\varepsilon,T}$. For the energy $\mathcal{E}_{2,T}$ (resp. $\mathcal{E}_{\nabla,T}$) the solution obtained has the smoothness described in Proposition 4.1 (resp. 4.4), that is $u^* \in W^{1,\infty}(O)$ and $w^* \in W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$ (resp. $u^* \in W^{1,2}(O) \cap W^{2,p}_{\mathrm{loc}}(O) \cap L^\infty(O)$ for any $p \in [1,\infty)$ and $w \in W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$).*

Let us point out that the convergence of the alternating optimization (Algorithm 1) holds also in the discrete domain.

Proof. Being similar, we give the details only for the case of the patch NL-Poisson energy.

Step 1. Basic estimates. Let us prove that

$$\kappa \sum_{k=0}^{N} \|w^{k+1} - w^k\|^2 + \mathcal{E}_{\nabla,T}(u^{N+1}, w^{N+1}) \leq \mathcal{E}_{\nabla,T}(u^0, w^0) \tag{6.3}$$

for some $\kappa > 0$ and $\{w^k(x,y)\}_k$ is uniformly bounded in $W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$.

Let $h(u, w; x, y) = w(x,y)\|p_u(x) - p_{\hat{u}}(y)\|^2_{g,\nabla} + Tw(x,y)\log w(x,y)$. We may write $\mathcal{E}_{\nabla,T}(u,w) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} h(u,w;x,y)\mathrm{d}y\mathrm{d}x$.

To prove (6.3), let us observe that

$$\begin{aligned} h(u^k, w^k; x, y) - h(u^k, w^{k+1}; x, y) = {} & \frac{\partial h}{\partial w}(u^k, w^{k+1}; x, y)(w^k(x,y) - w^{k+1}(x,y)) \\ & + \frac{\partial^2 h}{\partial w^2}(u^k, \overline{w}; x, y)(w^k(x,y) - w^{k+1}(x,y))^2 \end{aligned}$$

for an intermediate value $\overline{w}(x,y) \in [w^k(x,y), w^{k+1}(x,y)]$. Since $\frac{\partial h}{\partial w}(u^k, w^{k+1}) = 0$ and $\frac{\partial^2 h}{\partial w^2}(u^k, \overline{w}) = \frac{T}{\overline{w}}$, it suffices to prove that the sequence w^k is bounded independently of k, because this implies that also is $\overline{w}$. In that case we have

$$h(u^k, w^k; x, y) - h(u^k, w^{k+1}; x, y) \geq \kappa (w^k(x,y) - w^{k+1}(x,y))^2 \tag{6.4}$$

for some $\kappa > 0$.

To prove that w^k is bounded independently of k it suffices to observe that

$$w^k(x,y) = \frac{1}{Z_{\nabla,T}(u^{k-1}; x)} \exp\left(-\frac{1}{T} \|p_{u^{k-1}}(x) - p_{\hat{u}}(y)\|^2_{g,\nabla} \right), \tag{6.5}$$

where the normalizing factor $Z_{\nabla,T}(u^{k-1}; x)$ is given by

$$Z_{\nabla,T}(u^{k-1}; x) = \int_{\widetilde{O}^c} \exp\left(-\frac{1}{T} \|p_{u^{k-1}}(x) - p_{\hat{u}}(y)\|^2_{g,\nabla} \right) \mathrm{d}y. \tag{6.6}$$

Now, we observe that by Lemma 4.5, u^k is uniformly bounded in $W^{1,2}(O)$, and by Lemma 4.6 $\|p_{u^{k-1}}(x) - p_{\hat{u}}(y)\|^2_{g,\nabla}$ is uniformly bounded. This implies that there exist $b > a > 0$ independent of k such that

$$a \leq Z_{\nabla,T}(u^{k-1}; x) \leq b.$$

By the results of Section 4.2 this implies that $\{w^k(x,y)\}_k$ is uniformly bounded in $W^{1,\infty}(\widetilde{O} \times \widetilde{O}^c)$.

Now, using (6.4) we obtain

$$\begin{aligned}
\kappa \|w^k - w^{k+1}\|_2^2 &\leq \mathcal{E}_{\nabla,T}(u^k, w^k) - \mathcal{E}_{\nabla,T}(u^k, w^{k+1}) \\
&= \mathcal{E}_{\nabla,T}(u^k, w^k) - \mathcal{E}_{\nabla,T}(u^{k+1}, w^{k+1}) + \mathcal{E}_{\nabla,T}(u^{k+1}, w^{k+1}) - \mathcal{E}_{\nabla,T}(u^k, w^{k+1}) \\
&\leq \mathcal{E}_{\nabla,T}(u^k, w^k) - \mathcal{E}_{\nabla,T}(u^{k+1}, w^{k+1}),
\end{aligned}$$

since $\mathcal{E}_{\nabla,T}(u^{k+1}, w^{k+1}) - \mathcal{E}_{\nabla,T}(u^k, w^{k+1}) \leq 0$ because u^{k+1} is given by (6.2). By adding from $k = 0, \ldots, N$, we get (6.3).

Step 2. Convergence to a critical point of $\mathcal{E}_{\nabla,T}$ and regularity. By Step 1 we may extract a subsequence k_j such that w^{k_j} weakly converges to some w^* in L^p for all $p \in [1, \infty]$ and u^{k_j} converges to some $u^* \in W^{1,2}(O)$. By (6.3) also w^{k_j+1} converges to w^* in L^p for all $p \in [1, \infty]$.

The equations satisfied by u^{k+1}, w^{k+1} are

$$\begin{aligned}
\Delta u^{k+1}(z) &= \operatorname{div} \boldsymbol{v}^{k+1}(z) && z \in O \\
u^{k+1}(z) &= \hat{u}(z) && z \in \partial O,
\end{aligned} \tag{6.7}$$

where

$$\boldsymbol{v}^{k+1}(z) = \int_{\Omega_p} g(h) \int_{\widetilde{O}^c} w^{k+1}(z-h, y) \nabla \hat{u}(y+h) \mathrm{d}y \mathrm{d}h \tag{6.8}$$

and $w^{k+1}(x,y)$ is given by (6.5) and (6.6) with k replaced by $k+1$.

Notice that

$$v^{k_j}(z), v^{k_j+1}(z) \to v^*(z) := \int_{\Omega_p} g(h) \int_{\widetilde{O}^c} w^*(z-h, y) \nabla \hat{u}(y+h) \mathrm{d}y \mathrm{d}h$$

as $j \to \infty$. The convergence is also strong, since v^k is uniformly bounded. Then, using (6.7) we have that

$$\|\nabla u^k - \nabla u^{k+1}\|_2^2 = \int_{\mathbb{R}^N} (\boldsymbol{v}^k - \boldsymbol{v}^{k+1})(\nabla u^k - \nabla u^{k+1})\, dz \leq \|\boldsymbol{v}^k - \boldsymbol{v}^{k+1}\|_2 \|\nabla u^k - \nabla u^{k+1}\|_2.$$

Since both u^k and u^{k+1} have the same boundary values, $u^k - u^{k+1}$ converges to 0 in $W^{1,2}(O)$. Thus u^{k_j}, u^{k_j+1} both converge to u^* in $L^2(O)$. We have

$$w^*(x,y) = \frac{1}{Z_{\nabla,T}(u^*;x)} \exp\left(-\frac{1}{T}\|p_{u^*}(x) - p_{\hat{u}}(y)\|_{g,\nabla}^2\right), \tag{6.9}$$

where the normalizing factor $Z_{\nabla,T}(u^*;x)$ is given by

$$Z_{\nabla,T}(u^*;x) = \int_{\widetilde{O}^c} \exp\left(-\frac{1}{T}\|p_{u^*}(x) - p_{\hat{u}}(y)\|_{g,\nabla}^2\right) \mathrm{d}y. \tag{6.10}$$

Notice that u^* satisfies the boundary condition in (6.7). Thus $(u^*, w^*) \in \mathcal{A}_\nabla$ is a critical point of $\mathcal{E}_{\nabla,T}(u,w)$. Since $\mathcal{E}_{\nabla,T}(u^*,w)$ is a strictly convex function of w, then w^* is a minimum of $\mathcal{E}_{\nabla,T}(u^*,w)$. Since $\mathcal{E}_{\nabla,T}(u,w^*)$ is a strictly convex function of u, then u^* is a minimum of $\mathcal{E}_{\nabla,T}(u,w^*)$. By Remark 1, (u^*,w^*) we have the regularity stated in the statement. □

Let us notice that the iterations of the alternating optimization algorithm coincide with the Expectation Maximization algorithm (EM). Indeed, (6.1) is the E-step, while (6.2) is the M-step. Using the notation of Section 3, if

$$w_{\varepsilon,T}(u;x,y) = \frac{1}{Z_{\varepsilon,T}(u;x)} \exp\left(-\frac{1}{T}\|p_u(x) - p_{\hat{u}}(y)\|_{g,\varepsilon}^2\right), \tag{6.11}$$

where $Z_{\varepsilon,T}(u;x)$ is the corresponding normalization factor, then we may write

$$\frac{1}{T}\mathcal{E}_{\varepsilon,T}(u,w) = \int_{\widetilde{O}} \mathrm{KL}(w_{\varepsilon,T}(u;x,\cdot), w(x,\cdot))\, dx - \mathcal{L}_\varepsilon(u),$$

where

$$\mathcal{L}_\varepsilon(u) = \int_{\widetilde{O}} \log\, Z_{\varepsilon,T}(u;x)\, dx \tag{6.12}$$

corresponds to the so called *marginal likelihood* in the context of EM. The alternating optimization algorithm converges (modulo subsequences) to stationary points of $\mathcal{L}_\varepsilon(u)$. Notice that given u, the solution of $\min_w \mathcal{E}_{\varepsilon,T}(u,w)$ is given by $w_{\varepsilon,T}(u)$. Note that

$$-T\mathcal{L}_\varepsilon(u) = \mathcal{E}_{\varepsilon,T}(u, w_{\varepsilon,T}(u)) \leq \mathcal{E}_{\varepsilon,T}(u,w) \quad \forall (u,w)$$

and

$$\min_{(u,w)} \mathcal{E}_{\varepsilon,T}(u,w) = \min_u \min_w \mathcal{E}_{\varepsilon,T}(u,w) = \min_u \mathcal{E}_{\varepsilon,T}(u, w_{\varepsilon,T}(u)) = \min_u -T\mathcal{L}_\varepsilon(u).$$

Thus, functional $\mathcal{E}_{\varepsilon,T}(u,w)$ is equivalent to $-\mathcal{L}_\varepsilon(u)$ in the sense that both have the same minima. The alternating optimization algorithm converges to a critical point of both of them. More precisely:

PROPOSITION 6.2. *Any limit point of the sequence* $\{u^k\}_{k=0}^{\infty}$ *defined by Algorithm 1, is a stationary point* u^* *of* $\mathcal{L}_\varepsilon(u)$ *and* $\mathcal{L}_\varepsilon(u^k)$ *converges monotonically to* $\mathcal{L}_\varepsilon(u^*)$.

Proof. Indeed,

$$\begin{aligned}
&\mathcal{L}_\varepsilon(u^{k+1}) - \mathcal{L}_\varepsilon(u^k) = -\frac{1}{T}\mathcal{E}_{\varepsilon,T}(u^{k+1}, w_{\varepsilon,T}(u^k)) + \frac{1}{T}\mathcal{E}_{\varepsilon,T}(u^k, w_{\varepsilon,T}(u^k)) \\
&+ \int_{\widetilde{O}} \mathrm{KL}(w_{\varepsilon,T}(u^{k+1}; x, \cdot), w_{\varepsilon,T}(u^k; x, \cdot))\, dx - \int_{\widetilde{O}} \mathrm{KL}(w_{\varepsilon,T}(u^k; x, \cdot), w_{\varepsilon,T}(u^k; x, \cdot))\, dx \\
&\geq -\frac{1}{T}\mathcal{E}_{\varepsilon,T}(u^{k+1}, w_{\varepsilon,T}(u^k)) + \frac{1}{T}\mathcal{E}_{\varepsilon,T}(u^k, w_{\varepsilon,T}(u^k)) \geq 0.
\end{aligned}$$

Thus $\mathcal{L}_\varepsilon(u^k)$ is increasing. ∎

Let us notice that one can prove Proposition 6.2 by adapting the proof of Theorem 2 in [58].

REMARK 5. We can also interpret the above expressions in terms of a probabilistic model in the space of patches. The marginal likelihood (6.12) can be interpreted by noticing that $Z_{\varepsilon,T}(u; x)$ is a density estimate (in the patch space) of the set of patches in O^c: it corresponds to the total unnormalized similarity of patch $p_u(x)$.

The minimizer (u^*, w^*) are obtained when for all $(x, y) \in \widetilde{O} \times \widetilde{O}^c$, $w^*(x, y) = p_{\varepsilon,\beta}(u^*; x, y)$ (*i.e.* normalized Gaussian weights), and the patches of the inpainted image are in regions of high density in the patch space. This provides a geometric intuitive interpretation of our variational formulation. The image is considered as an ensemble of overlapping patches. Known patches in $\widetilde{O}^c$ are fixed, forming a patch density model used to estimate the patches in $\widetilde{O}$.

6.1. The relaxed correspondence model (discrete case). The translation of the previous approach to the relaxed correspondence model $\mathcal{E}_{\varepsilon,0}$ described in Section 5 has to face in its continuous version the difficulty of the lack of estimates ensuring the compactness of the iterates of (6.13) and (6.14) in Algorithm 2. In the discrete case, the convergence can be proved thanks to the convexity of $\mathcal{E}_{\varepsilon,0}(u, \nu)$ in each variable when the other is fixed. Thus, we work only in the discrete case.

Algorithm 2 Alternate minimization of $\mathcal{E}_{2,0}$.

Initialization: choose u^0 with $\|u^0\|_\infty \leq \|\hat{u}\|_\infty$.

For each $k \in \mathbb{N}$ solve

$$\nu^k = \operatorname*{arg\,min}_{\nu \in \mathcal{MP}} \mathcal{E}_{\varepsilon,0}(u^k, \nu) \tag{6.13}$$

$$u^{k+1} = \operatorname*{arg\,min}_{u} \mathcal{E}_{\varepsilon,0}(u, \nu^k) \tag{6.14}$$

PROPOSITION 6.3. *There exists a subsequence which converges to a critical point of the energy* $\mathcal{E}_{\varepsilon,0}$.

Proof. Since u^k and ν^k are bounded, there is a subsequence (u^{k_j}, ν^{k_j}) converging to $(\bar{u}, \bar{\nu})$. Notice that if ν is fixed, then the solution of $\min_u \mathcal{E}_{\varepsilon,0}(u, \nu)$ is unique. Clearly $\bar{\nu}$ is a minimum of $\nu \to \mathcal{E}_{\varepsilon,0}(\bar{u}, \nu)$. Proceeding as in [13], one can prove that $u^{k_j+1} - u^{k_j}$ converges to 0 and deduce that $\bar{u}$ is a minimum of $u \to \mathcal{E}_{\varepsilon,0}(u, \bar{\nu})$. ∎

The convergence of the first step of the algorithm (6.13) is the object of next

Section. The solution of (6.14) is explicitly given by

$$\bar{u}(x) = \sum_{\mathbb{Z}^2} g * (\chi_{\widetilde{O}} \chi_{\widetilde{O}^c} \nu^k)(x - \cdot, y - \cdot)\hat{u}(y) \tag{6.15}$$

for the energy $\mathcal{E}_{2,0}$ and requires the solution of a discrete version of Poisson equation for $\mathcal{E}_{\nabla,0}$.

7. Computation of the Nearest Neighbor Field. In this Section we discuss some aspects on the numerical minimization of the inpainting functionals. Throughout the Section we consider discretized versions of the inpainting domain and its complement, $\widetilde{O} = \widetilde{O} \cap \mathbb{Z}^2$ and $\widetilde{O}^c = \widetilde{O}^c \cap \mathbb{Z}^2$. To avoid a cumbersome notation, we slightly modify it in this Section (for instance some arguments of functions will be denoted as subindices).

For minimizing the functionals derived from (2.1), (2.2) we use alternating optimization schemes (Algorithms 1 and 2). Most of the computational load is caused by the updating of the weights. In this Section we discuss the convergence properties of PatchMatch, an algorithm recently introduced by Barnes *et al.* [9], which we use to speed-up the computation of the similarity weights. For other aspects of the numerical implementation we refer to Section 8. More details can be found in [5].

For $T > 0$, the computation of the weight function w is of order $O(|\widetilde{O}||\widetilde{O}^c||\Omega_p|)$. This is also the case in the limit $T = 0$, namely for $\mathcal{E}_{\varepsilon,0}$. In that case, as shown in Proposition 5.3, there are minima given by measurable measure-valued maps ν determined by a measurable correspondence map $\varphi : \widetilde{O} \to \widetilde{O}^c$. This allows us to express the energy directly in terms of the unknown map φ, instead of the measure-valued map ν. Thus, when considering the optimization of $\mathcal{E}_{\varepsilon,0}$ the weights update step is substituted by a minimization w.r.t. a correspondence map φ,

$$\varphi_x \in \operatorname*{arg\,min}_{\xi \in \mathbb{Z}^2} U_x(\xi), \quad \text{for all } x \in \widetilde{O},$$

where the energy U_x corresponds to the patch error function

$$U_x(\xi) = \begin{cases} \varepsilon(p_u(x) - p_{\hat{u}}(\xi)) & \text{if } \xi \in \widetilde{O}^c \\ +\infty & \text{otherwise.} \end{cases}$$

Although the patch error does not have to be a metric, we will refer to $p_{\hat{u}}(\varphi_x)$ as the *nearest patch* or *nearest neighbor* of $p_u(x)$. Following [9], we denote the correspondence map $\varphi : \widetilde{O} \to \widetilde{O}^c$ as the *nearest neighbor field* (NNF). A brute force search for the NNF also conveys $O(|\widetilde{O}||\widetilde{O}^c||\Omega_p|)$ operations.

PatchMatch is a very efficient algorithm for approximating the NNF [9]. The search for the nearest neighbor is performed simultaneously over the points in $\widetilde{O}$ based in the following heuristic: since query patches overlap, the *offset* $\varphi_x - x$ of a good match at x is likely to lead to a good match for the adjacent points of x as well. It is an iterative algorithm which starting from a random initialization, alternates between steps of propagation of good offsets and random search.

For describing them, we need some definitions. Pixels in $\widetilde{O} = \{x_1, x_2, \ldots, x_{|\widetilde{O}|}\}$ are sorted according to the lexicographical order in $\mathbb{Z}^2$. For any $x \in \widetilde{O}$, let $\mathcal{N}_4(x) = \{z \in \widetilde{O} : 0 < |z - x| \leqslant 1\}$ be its 4-neighborhood. We consider a transition probability kernel $Q : \widetilde{O}^c \times \mathcal{B} \to [0, 1]$, where $\mathcal{B}$ is a σ-algebra in $\widetilde{O}^c$ (the subsets of $\widetilde{O}^c$ in our

discrete case). Finally, let us define the notation,

$$\eta \wedge_x \xi = \begin{cases} \eta & \text{if } U_x(\eta) \leq U_x(\xi) \\ \xi & \text{if } U_x(\eta) > U_x(\xi), \end{cases}$$

where $\xi, \eta \in \widetilde{O}^c$.

The PatchMatch algorithm is described in Algorithm 3. The computational order is $O(|\Omega_p||\widetilde{O}|)$, whereas the memory requirements are of $O(|\widetilde{O}|)$. For most applications, a few iterations after a random initialization are often sufficient. Our implementation, coded in C without optimizations, running on a notebook's dual core 1.8GHz CPU, takes between $3s$ to $4s$ to compute the correspondences for an image of 300×255, with a mask of $|\widetilde{O}| \approx 14.000$ pixels and $|\widetilde{O}^c| \approx 50.000$ pixels and a patch of size 7×7.

This algorithm can be extended to store queues of L offsets in an L-Nearest Neighbors Field [10] (see also [5]). This allows its application to the case $T > 0$, by truncating the support of $w(x, \cdot)$ to the L-nearest neighbors of $p_u(x)$. This increases the computational cost and the memory requirements by a factor of L. We note however, that although each iteration is more costly, the use of queues usually reduces the required number of iterations.

Algorithm 3 PatchMatch with propagation of offsets.

Initialization. Choose $\varphi_x^0 \sim \mathcal{U}(\widetilde{O}^c)$, i.e., randomly with a uniform distribution.

For each $n \in \mathbb{N}$,

Random search. For each $x \in \widetilde{O}$ draw $S\varphi_x^n \sim Q(\varphi_x^n, \cdot)$. Set $\varphi_x^{n+\frac{1}{2}} = \varphi_x^n \wedge_x S\varphi_x^n$.

Forward propagation. If n is odd, for each $i = 1, \ldots, |\widetilde{O}|$, set

$$\varphi_{x_i}^{n+1} = \varphi_{x_i}^{n+\frac{1}{2}} \wedge_{x_i} \left(\bigwedge_{x_i} \{\varphi_{x_j}^{n+1} + x_i - x_j : x_j \in \mathcal{N}_4(x_i), j < i\} \right).$$

If n is even, invert the direction of propagation (*backward propagation*).

7.1. Convergence of the PatchMatch algorithm. In this Section we discuss the convergence properties of the PatchMatch algorithm (for $L = 1$, *i.e.* without considering queues). For simplicity we will prove the convergence result for a different version of the PatchMatch algorithm. The difference lies in the propagation step: In the original version of Barnes *et al.* [9], a node z will propagate the offset $\varphi_z - z$ to x. In our simplified version, the absolute position $\varphi_z \in \widetilde{O}^c$ is propagated instead (Algorithm 4). The arguments for the simplified "position-propagation" case can be applied to the "offset-propagation" case (with more involved computations). The latter is much more relevant from the practical point of view. This is briefly discussed in Section 7.1.1. Before proceeding to the convergence result, it is necessary to add some additional structure.

We will consider that elements in $\widetilde{O}$ correspond to the vertices of a *directed acyclic graph* (DAG) $G = (\widetilde{O}, E)$, where $E \subset \widetilde{O} \times \widetilde{O}$ denotes the edge set. We define the edge set as follows:

$$E = \{(x, y) \in \widetilde{O} \times \widetilde{O} : y \in \mathcal{N}_4(x), x < y\}.$$

By $x < y$ we mean that x precedes y in the lexicographical order. Note that the lexicographical order is a topological order for the resulting DAG. Paths in G will be denoted by $c = (c_1, \dots, c_{n_c}) \in \widetilde{O}^{n_c}$, where $n_c \in \mathbb{N}$ is the length of the path. Given any pair of nodes, $x, z \in \widetilde{O}$, we will denote by $\mathcal{P}(z, x)$ the set of paths from z to x. A node $z \in \widetilde{O}$ is said to be an *ancestor* of x if $\mathcal{P}(z, x) \neq \emptyset$. Note that if z is an ancestor of x, then z comes before x in the lexicographical ordering. Similarly, z is a *descendant* of x if $\mathcal{P}(x, z) \neq \emptyset$ (*i.e.* x is an ancestor of z). We will write $\mathcal{A}(x)$ and $\mathcal{D}(x)$ for the set of ancestors and descendants of node x, respectively.

Algorithm 4 Propagation of positions.

Given $n \in \mathbb{N}$, and $\varphi^{n+\frac{1}{2}}$:
Forward propagation. If n is odd, for each $i = 1, \dots, |\widetilde{O}|$, set

$$\varphi_{x_i}^{n+1} = \varphi_{x_i}^{n+\frac{1}{2}} \wedge_{x_i} \left(\bigwedge_{x_i} \{ \varphi_{x_j}^{n+1} : x_j \in \mathcal{N}_4(x_i), j < i \} \right).$$

If n is even, invert the direction of propagation (*backward propagation*).

The following proposition provides a bound on the convergence rate (in probability) for Algorithm 4. Without loss of generality we will assume throughout this Section that $\min_\xi U_x(\xi) = 0$ for all $x \in \widetilde{O}$.

PROPOSITION 7.1. *Assume that for each pair $x, y \in \widetilde{O}$, we have that $d_{x,y} := \|U_x - U_y\|_\infty < +\infty$. Assume that $\widetilde{O}^c$ is compact (and therefore finite) and that $Q(x, A) > 0$, for all $x \in \widetilde{O}$, $A \subset R$. Then, the sequence (φ^n) defined by the PatchMatch algorithm converges to a minimizer of the total energy U, in the sense that*

$$\lim_{n \to \infty} P(U_x(\varphi_x^n) > \epsilon) = 0, \qquad \text{for all } \epsilon > 0, x \in \widetilde{O}.$$

Moreover, we have that

$$P(U_x(\varphi_x^{n+1}) > \epsilon) \leqslant \prod_{z \in \mathcal{A}(x)} C(z, \epsilon - \ell_{z,x}) P(U_x(\varphi_x^n) > \epsilon), \tag{7.1}$$

where $\ell_{z,x}$ is the length of the minimal path from z to x:

$$\ell_{z,x} := \begin{cases} \min_{c \in \mathcal{P}(z,x)} \sum_{i=2}^{n_c} d_{c_{i-1}, c_i} & \text{if } \mathcal{P}(z, x), \neq \emptyset, \\ +\infty, & \text{if } \mathcal{P}(z, x) = \emptyset, \end{cases}$$

and for each $z \in \widetilde{O}$, $C(z, \cdot) : \mathbb{R} \to [0, 1]$ is a non-increasing function defined by:

$$C(z, a) := \sup_{\eta \in \{U_z > a\}} Q(\eta, \{U_z > a\}).$$

For $a > 0$, $C(z, a) < 1$.

The proof of Proposition 7.1 relies on the following lemma.

LEMMA 7.2. *Assume that for each pair $x, y \in \widetilde{O}$, we have that $d_{x,y} := \|U_x - U_y\|_\infty < +\infty$. Let us consider an assignment φ resulting from a propagation step. Then we have that for each pair of nodes $x, z \in \widetilde{O}$,*

$$U_x(\varphi_x) > \epsilon \quad \Rightarrow \quad U_z(\varphi_z) > \epsilon - \ell_{z,x}.$$

Proof. Let us consider a path $c \in \mathcal{P}(z,x)$. We have that for any $a > 0$,

$$U_{c_i}(\varphi_{c_i}) > a \quad \Rightarrow \quad U_{c_i}(\varphi_{c_{i-1}}) > a$$

(otherwise, $\varphi_{c_{i-1}}$ would have been *propagated* to node c_i). Since $\|U_{c_i} - U_{c_{i-1}}\|_\infty = d_{c_{i-1},c_i}$, we have that

$$U_{c_i}(\varphi_{c_i}) > a \quad \Rightarrow \quad U_{c_{i-1}}(\varphi_{c_{i-1}}) > a - d_{c_{i-1},c_i}.$$

A simple recursion results in

$$U_x(\varphi_x) > \epsilon \quad \Rightarrow \quad U_z(\varphi_z) > \epsilon - \sum_{i=2}^{n_c} d_{c_{i-1},c_i}.$$

Thus each path from z to x imposes a bound over $U_z(\varphi_z)$. The intersection of all of them is given by $U_z(\varphi_z) > \epsilon - \ell_{z,x}$. □

We now prove Proposition 7.1.

Proof. Let us consider $x \in \widetilde{O}$. Since φ^{n+1} is the result of a search step followed by a propagation step, we can apply Lemma 7.2. Thus,

$$\begin{aligned} U_x(\varphi_x^{n+1}) > \epsilon \quad &\Rightarrow \quad U_z(\varphi_z^{n+1}) > \epsilon - \ell_{z,x}, \quad \forall z \in \widetilde{O}, \\ &\Rightarrow \quad U_z(\varphi_z^{n}) > \epsilon - \ell_{z,x} \text{ and } U_z(S\varphi_z^{n}) > \epsilon - \ell_{z,x}, \quad \forall z \in \widetilde{O}, \end{aligned}$$

where the last implication is due to the random search step. Taking probabilities we have that

$$\begin{aligned} &P(U_x(\varphi_x^{n+1}) > \epsilon) \leqslant P(U_z(\varphi_z^n) > \epsilon - \ell_{z,x} \text{ and } U_z(S\varphi_z^n) > \epsilon - \ell_{z,x}, \forall z \in \widetilde{O}) \\ &= \prod_{z \in \widetilde{O}} P(U_z(S\varphi_z^n) > \epsilon - \ell_{z,x} | U_z(\varphi_z^n) > \epsilon - \ell_{z,x}) P(U_z(\varphi_z^n) > \epsilon - \ell_{z,x}, \forall z \in \widetilde{O}) \\ &\leqslant \prod_{z \in \widetilde{O}} P(U_z(S\varphi_z^n) > \epsilon - \ell_{z,x} | U_z(\varphi_z^n) > \epsilon - \ell_{z,x}) P(U_x(\varphi_x^n) > \epsilon). \end{aligned}$$

The second equality is due to Bayes' formula for conditional probabilities, and to the conditional independence of the random searches given φ^n. In the latter inequality we have used that $P(U_z(\varphi_z^n) > \epsilon - \ell_{z,x}, \forall z \in \widetilde{O}) \leqslant P(U_x(\varphi_x^n) > \epsilon)$ since the l.h.s. corresponds to the probability of the intersection of several events, while the r.h.s. is the probability of only one of such events.

Given $z \in \widetilde{O}$, we denote by $P^{z,n}$ the probability distribution of φ_z^n (*i.e.* $P(\varphi_z^n \in A) = P^{z,n}(A)$, for $A \in \mathcal{B}(\widetilde{O}^c)$). Then, by the definition of conditional probability, we have:

$$P(U_z(S\varphi_z^n) > a | U_z(\varphi_z^n) > a) = \frac{1}{P^{z,n}(\{U_z > a\})} \int_{U_z(\eta) > a} P^{z,n}(\mathrm{d}\eta) \int_{U_z(\xi) > a} Q(\eta, \mathrm{d}\xi).$$

Notice that $P_a^{z,n} := [P^{z,n}(\{U_z > a\})]^{-1} P^{z,n}$ is a probability when restricted to the upper level set $\{U_z > a\}$. Therefore, the following bound holds

$$P(U_z(S\varphi_z^n) > a | U_z(\varphi_z^n) > a) \leqslant \sup_{\eta \in \{U_z > a\}} Q(\eta, \{U_z > a\}) =: C(z,a).$$

The coefficient $C(z,a)$ a supremum of probabilities, thus $C(z,a) \in [0,1]$. The function $C(z,\cdot) : \mathbb{R} \to [0,1]$ is non-increasing, since for $a_1, a_2 \in \mathbb{R}$, $a_1 > a_2$ we have that $\{U_z > a_1\} \subseteq \{U_z > a_2\}$. Finally, since U_z is non-negative with a minimum value of 0, for $a > 0$, we have that $Q(\eta, \{U_z > a\}) < 1$, for all $\eta \in \widetilde{O}^c$. Since $\widetilde{O}^c$ is compact (and therefore finite in the discrete setting) we have that $C(z,a) < 1$, for $a > 0$. ∎

REMARK 6. The efficiency of the PatchMatch is mostly given by the propagation steps, when nodes collaborate by sharing their findings. This is reflected by (7.1). For comparison, consider a PatchMatch algorithm without propagation. Each φ_x is searched for independently for each $x \in \widetilde{O}$. In that case, the bound on the rate of convergence (7.1) reduces to

$$P(U_x(\varphi_x^{n+1}) > \epsilon) \leqslant C(x,\epsilon) P(U_x(\varphi_x^n) > \epsilon).$$

The speed-up given by the propagation corresponds to $\prod_{\substack{z \in \mathcal{A}(x) \\ z \neq x}} C(z, \epsilon - \ell_{z,x})$. Note that only those $z \in \mathcal{A}(x)$ with $\ell_{z,x} < \epsilon$ contribute to lower the bound.

7.1.1. Registered propagation. Let us now address the propagation of offsets (Algorithm 3). We extend the definition of the energies to $\mathbb{Z}^2$ by taking $U_x(\xi) = +\infty$ if $\xi \in \mathbb{Z}^2 \setminus \widetilde{O}^c$. For each pair of connected nodes $(z,x) \in E$ we are given a transformation $T_{x,z} : \widetilde{O}^c \to \mathbb{Z}^2$ (which is $T_{x,z}(\xi) = \xi - z + x$ in Algorithm 3). The *registered propagation* is defined as follows:

$$\varphi_{x_i}^{n+1} = \varphi_{x_i}^{n+\frac{1}{2}} \wedge_{x_i} \left(\bigwedge_{x_i} \{ T_{x_i,z}(\varphi_z^{n+1}) : (z,x_i) \in E \} \right).$$

With analogous computations (slightly more involved), one can prove that this version of the PatchMatch also converges, with the following bound for the rate of convergence:

$$P(U_x(\varphi_x^{n+1}) > \epsilon) \leqslant \prod_{z \in \mathcal{A}(x)} C(z, \beta(T,z,x,\epsilon)) P(U_x(\varphi_x^n) > \epsilon), \tag{7.2}$$

where

$$\beta(T,z,x,\epsilon) = \max_{c \in \mathcal{P}(z,x)} \min_{i=1,\dots,n_c} \left\{ U^*(T,c,\epsilon)_i - \sum_{j=i+1}^{n_c} d(T)_{c_{j-1},c_j} \right\}, \tag{7.3a}$$

$$U^*(T,c,\epsilon)_i = \begin{cases} \epsilon & \text{if } i = 1, \\ \min\{U_{c_{i-1}}(\xi) : \xi \in \widetilde{O}^c \setminus T_{c_i,c_{i-1}}^{-1}(\widetilde{O}^c)\} & \text{if } i = 2,\dots,n_c, \end{cases} \tag{7.3b}$$

$$d(T)_{x,z} = \|U_x \circ T_{x,z} - U_z\|_{\infty, T_{x,z}^{-1}(\widetilde{O}^c)}. \tag{7.3c}$$

This is a consequence of the fact that we are allowing the transformations $T_{x,z}$ to map some $\xi \in \widetilde{O}^c$ outside $\widetilde{O}^c$, *i.e.* $T_{x,z}(\xi) \in \mathbb{Z}^2 \setminus \widetilde{O}^c$. If $T_{x,z}(\varphi_z) \in \mathbb{Z}^2 \setminus \widetilde{O}^c$, then it is not propagated to node x. Thus, the fact that $U_x(\varphi_x) > \epsilon$, implies that either $\varphi_z \in T_{x,z}^{-1}(\widetilde{O}^c)$ and $U_z(\varphi_z) > \epsilon - d(T)_{x,z}$, or $\varphi_z \in \widetilde{O}^c \setminus T_{x,z}^{-1}(\widetilde{O}^c)$, in which case, we do not have control over $U_z(\varphi_z)$ other than

$$U_z(\varphi_z) > \min\{U_z(\xi) : \xi \in \widetilde{O}^c \setminus T_{x,z}^{-1}(\widetilde{O}^c)\}.$$

This gives rise to the U^* coefficients in (7.2).

Observe that in the case in which $T_{x,z}(\widetilde{O}^c) \subset \widetilde{O}^c$, for all $(z,x) \in E$, then one recovers

$$\beta(T,z,x,\epsilon) = \epsilon - \ell(T)_{z,x} = \epsilon - \min_{c\in\mathcal{P}(z,x)} \sum_{i=2}^{n_c} d(T)_{c_{i-1},c_i},$$

as in (7.1). The transformations $T_{x,z}$ should be chosen to lower the *edge cost* $d(T)_{x,z}$, by *registering* U_x and U_z.

In [9] the propagation is performed with a transformation $T_{x,z}(\xi) = \xi - z + x$, which corresponds to the propagation of the offset $\xi - z$. Let us estimate the corresponding bound according to (7.2). We denote by $e_0 = (0,1)$ and $e_1 = (1,0)$. The parents of node $x \in \widetilde{O}$ are $x - e_i$ with $i = 0,1$, and correspondingly $T_{x,x-e_i}(\xi) = \xi + e_i$. Note that $\widetilde{O}^c \setminus T^{-1}_{x,x-e_i}(\widetilde{O}^c) = \widetilde{O}^c \setminus (\widetilde{O}^c - e_i)$. To simplify the discussion, we assume that

$$\min\{U_z(\xi) : \xi \in \widetilde{O}^c \setminus (\widetilde{O}^c - e_i)\} > \kappa > 0$$

for all $z \in \widetilde{O}$. Then, as can be seen from (7.3a) and (7.3b), for all $\epsilon < \kappa$, we have that $\beta(T,z,x,\epsilon) = \epsilon - \min_{c\in\mathcal{P}(z,x)} \sum_{i=2}^{n_c} d(T)_{c_{i-1},c_i}$.

For $d(T)_{x,x-e_i}$ we have

$$\begin{aligned} d(T)_{x,x-e_i} &= \|U_x T_{x,x-e_i} - U_{x-e_i}\|_{\infty, T^{-1}_{x,x-e_i}} \\ &= \sup_{\xi\in T^{-1}_{x,x-e_i}(\widetilde{O}^c)} |g * \mathrm{e}(u(T_{x,x-e_i}(x-e_i)+\cdot) - \hat{u}(T_{x,x-e_i}(\xi)+\cdot)) - \\ &\qquad\qquad g * \mathrm{e}(u(x-e_i+\cdot) - \hat{u}(\xi+\cdot))| \\ &\leqslant \sup_{\xi\in T^{-1}_{x,x-e_i}(\widetilde{O}^c)} |(g\circ T_{x,x-e_i} - g)| * \mathrm{e}(u(x-e_i+\cdot) - \hat{u}(\xi+\cdot)). \end{aligned} \tag{7.4}$$

We have used that $x = T_{x,x-e_i}(x-e_i)$ and that if T is a translation $g * (f \circ T) = (g \circ T) * f$. The bound 7.4 corresponds to a patch error weighted by the kernel

$$\partial_i^+ g(h) := g \circ T_{x,x-e_i}(h) - g(h) = g(h+e_i) - g(h) \approx \partial_i g(h),$$

which is an approximation of the partial derivative of g. This is an interesting property, because if g is smooth, $|\partial_i^+ g(h)|$ is small (recall that we should minimize $d(T)$ to maximize the coefficients β). This is essentially the Lipschitz estimate of Lemma 4.2 in the present context.

The "propagation of offsets" exploits the overlap of neighboring patches in the image domain, suggesting that each node should be connected with its neighbors on the image grid. This supports the intuitions in [9].

7.2. Case $T > 0$: MCMC à la PatchMatch. As discussed before, a naive computation of w is of order $O(|\Omega_p||\widetilde{O}||\widetilde{O}^c|)$, which makes the inpainting algorithm of little practical use. The PatchMatch scheme allows an efficient computation of the correspondence map φ, when $T = 0$. But when $T > 0$, the nearest neighbor is only the mode of $w(x,\cdot)$. One way to circumvent this problem is to use an extension of the PatchMatch algorithm which finds the L nearest neighbors for each patch. For L large enough, most of the mass of $w(x,\cdot)$ will be captured by the L nearest neighbors.

This works for low values of T, but will yield a bad approximation when T is high. Furthermore, even when T is low, it is hard to set the appropriate value of L.

A more accurate estimation of w can be obtained with a Monte Carlo method, which consists of approximatin $w(x,\cdot)$ by a set of L samples $y_{x,1},\dots,y_{x,L}$ drawn independently from $w(x,\cdot)$. This opens an interesting problem, namely how to sample efficiently from a family of spatially indexed probability distributions, varying smoothly w.r.t. to the spatial index. The PatchMatch algorithm is designed to solve a related problem: finding the modes of this family of distributions. Thus, it can be expected that some sort of spatial collaboration between neighboring nodes (as the propagation step in PacthMatch) can yield a more efficient sampler. The investigation of these ideas in the context of MCMC will be the object of future research.

8. Experiments. We display some experiments using the energies above, namely the patch NL-means model, the patch NL-Poisson model, and a mixture of them (see Section 8.2). For inpainting real images we use the models with $T = 0$, but we also display an example obtained using the annealing scheme. Let us denote $\mathcal{E}(u,w)$ any of the energies used below. The numerical algorithm we use is based on the alternating optimization scheme (or EM scheme), Algorithms 1 and 2. The solution of the image update step is given explicitly as a non-local average (patch NL-means) or it requires the solution of a Poisson type equation (patch NL-Poisson and the combination of both). For computing the weights, we use the PatchMatch algorithm in case $T = 0$ or, for $T > 0$, an extension of it adapted to store queues of L offsets in an L-Nearest Neighbors Field [10] (see also [5]).

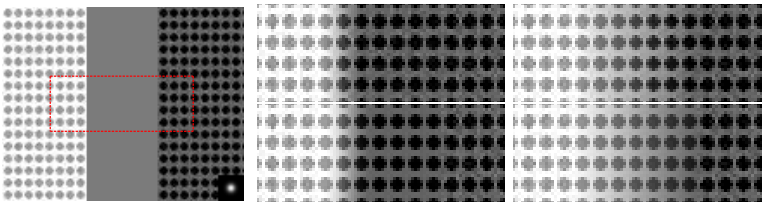

Figure 8.1: **Inpainting of a synthetic texture.** The initial condition is shown in the first column. The other four columns show a zoom (region in the red rectangle) of the results of patch NL-means, and -Poisson. Top row, $T = 0$, bottom row $T = 200$ and $T = 400$, respectively. The intra-patch weight kernel g is shown in the bottom right corner of the initial condition, it has a standard deviation $a = 5$ and the patch size is $s = 15$.

8.1. Effect of the selectivity parameter. First we consider the inpainting of a regular texture (shown in Figure 8.1) with two different mean intensities. The inpainting domain hides all patches on the boundary between the dark and bright textures. With this example we can test the ability of each method to *create* an interface between both regions. Situations like these are common in real inpainting problems, for instance due to inhomogeneous lighting conditions. We have also added Gaussian noise with standard deviation $\sigma = 10$ to show the influence of the selectivity parameter T. Each column of Figure 8.1 shows the results of the four methods described in the previous Section. We have tested each method with $T = 0$ (top row), and $T > 0$ (bottom row), chosen approximately to match the expected deviation of each patch error due to the presence of noise. The intra-patch weight kernel g is

shown in the bottom right corner of the initial condition. It is a Gaussian kernel with standard deviation $a = 5$ truncated to a square patch of size is $s = 15$.

Notice that the gradient-based method yields a much smoother shading of the texture. This is due to the fact that the image update step is computed as the solution of a PDE which diffuses the intensity values present at the boundary of the inpainting domain.

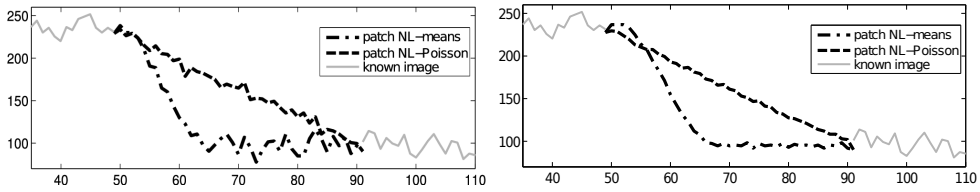

Figure 8.2: **Profiles of the results in Figure 8.1.** The profiles are taken from an horizontal line going between the circles in Figure 8.1. Top: results with $T = 0$ and bottom: results with $T > 0$.

As expected, the results using a higher value of T show some denoising. This effect can be better appreciated in the profiles shown in Figure 8.2, which depict the image values for a horizontal line between the circles. In the usual context of inpainting, in which the available data is not perturbed by noise, this denoising translates into an undesirable loss of texture quality (some details are treated as noise, a common effect in image enhancement). For that reason for the rest of the inpainting experiments shown in this paper we will consider $T = 0$. In other applications such as denoising or image regularization, the case of $T > 0$ becomes relevant. Although we do not pursue them in the present work, it would be interesting to explore the application of this formalism to more general settings following the line of [48] and our work in [28] on the reconstruction of of sparsely sampled images.

8.2. Results on real images. We present some results to illustrate how the inpainting functionals work on real images. We begin by briefly enumerating some implementation issues. For a detailed presentation of these issues we refer the reader to [5], where they will also find more results and a comparison with other state-of-the-art methods.

We consider two inpainting methods, variations of studied framework, namely patch NL-means, and -Poisson. As discussed in the preceding Section we set $T = 0$ to prevent from blurring.

Combination of gradients and intensities. For patch NL-Poisson the patch similarity weights w are computed based only on the gradients of the image. In most cases however, the gradient is not a good feature for measuring the patch similarity, and it is convenient to consider also the gray level/color data. For this reason we consider a convex combination of the patch error functions of patch NL-means and -Poisson:

$$\mathcal{E}_{2,\lambda,0}(u,w) = \int_{\widetilde{O}} \int_{\widetilde{O}^c} w(x,y) \left(\lambda \|p_u(x) - p_{\hat{u}}(y)\|_g^2 + (1-\lambda) \|p_u(x) - p_{\hat{u}}(y)\|_{\nabla,g}^2 \right) \mathrm{d}y \mathrm{d}x, \tag{8.1}$$

where the parameter $\lambda \in [0, 1)$ controls the mixture. The theoretical results of the paper hold in this case, under the assumptions of the patch NL-Poisson method. From

now on, we will use the term patch NL-Poisson to refer to this model. The case $\lambda = 1$ corresponds to the patch NL-means and will be considered separately. Typically we will set $\lambda \leqslant 0.1$, in this way we include some intensity information in the computation of the weights, without departing too much from the Poisson equation (2.6).

In this case the Euler equation w.r.t. u becomes:

$$\begin{aligned}&(1-\lambda)\Delta u(z) - \lambda u(z) = (1-\lambda)\mathrm{div}\,\boldsymbol{v}(z) - \lambda f(z), \quad \text{for } z \in O,\\ &u(z) = \hat{u}(z), \quad \text{for } z \in \partial O.\end{aligned}$$

Here

$$\begin{aligned}f(z) &:= \int_{\mathbb{R}^2} g * w_{\widetilde{O},\widetilde{O}^c}(z-\cdot, z'-\cdot)\hat{u}(z')\mathrm{d}z',\\ \boldsymbol{v}(z) &= \int_{\mathbb{R}^2} g * w_{\widetilde{O},\widetilde{O}^c}(z-\cdot, z'-\cdot)\nabla\hat{u}(z')\mathrm{d}z'.\end{aligned}$$

Observe that f corresponds to a patch NL-means image update, and $\boldsymbol{v}$ is a non-local average of the gradients in O^c. The problem is linear and can be solved for instance with a conjugate gradient scheme.

Multiscale scheme. Exemplar-based inpainting methods have a critical dependence with the size of the patch. Furthermore, when the inpainting domain is large in comparison with the patch, the energies have many local minima, and not all of them are *good* inpaintings. It is the common practice in the literature (*e.g.* [37, 39, 57]), to incorporate a multiscale scheme. It consists on applying sequentially the inpainting method on a Gaussian image pyramid, starting at the coarsest scale. The result at each scale is upsampled and used as initialization for the next finer scale. The patch size is constant through scales, which amounts (approximately) to minimize a sequence of energies with decreasing patch sizes without subsampling the image. This alleviates the critical dependence w.r.t. the size of the patch, helps in avoiding local minima and alleviates the computational cost. In our experiments, the size of the coarsest scale is a $10-20\%$ of the original size, except for a few cases which required less subsampling. The number of scales is set such that the subsampling rate is approximately 0.8 as in [57].

The results are shown in Figures 8.3, 8.4 and 8.5, classified according to the nature of the inpainting problem. The most important parameters are the patch size, the size of the coarsest scale, and λ. For almost all experiments we used patches of size s between 3×3 and 9×9. We used constant intra-patch weights ($g_a = 1/|\Omega_p|$). For the mixing coefficient λ for the patch NL-Poisson model in (8.1) we tested two configurations: $\lambda = 0.01$ and $\lambda = 0.1$. Recall that lower values of λ give a higher weight to the gradient component of the energy. This is appropriate for structured images with strong edges.

Copy regions and transition bands. Let us focus now on the solution of the functional in the limit case when $T \to 0$. In this case, the weights w are replaced by a map $\varphi : \widetilde{O} \to \widetilde{O}^c$. As a result of the alternating minimization scheme, $\varphi(x)$ corresponds to the center of the most similar patch to $p_u(x)$ (the nearest neighbor).

In Figure 8.7 we show some steps of the minimization process for the patch NL-means applied to a natural texture. The red curves on the bottom row depict the boundaries of the regions with constant offset with respect to the nearest neighbor. This offset is given by $t(x) = \varphi(x) - x$, for $x \in \widetilde{O}$. Since φ is the nearest neighbor field (NNF), we will refer to t as the *offset-to-the-nearest-neighbor field* (ONNF). The

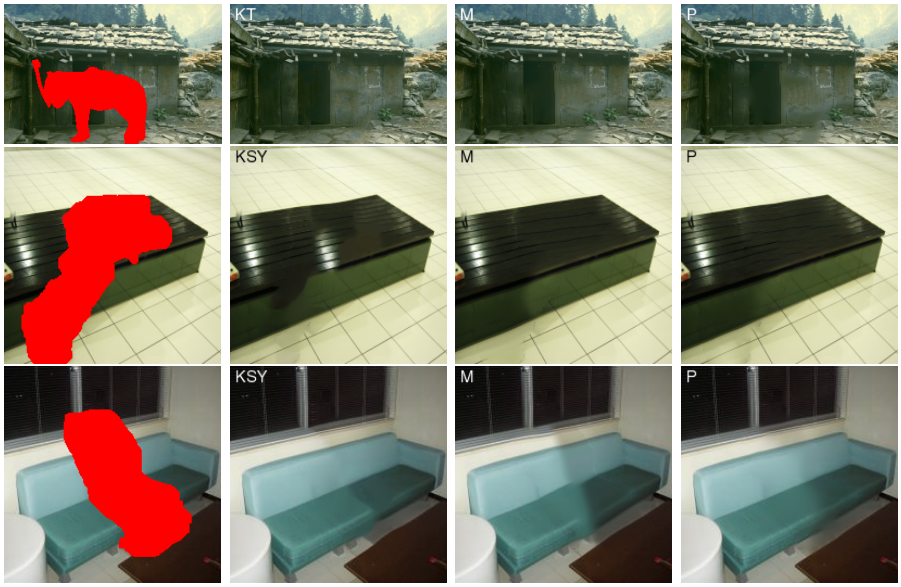

Figure 8.3: **Results on structured images.** KT: method of Komodakis and Tziritas [39] and KSY: method of Kawai *et al.* [37]. M, and P stand for patch NL-means, and -Poisson. From top to bottom *cabin*, *station bench* and *sofa*.

Figure 8.4: **Results on random textures.** KT: method of Komodakis and Tziritas [39]. M and P stand for patch NL-means and -Poisson. From top to bottom *baseball*, *bridge* and *golf*.

minimization process starts from a highly complex ONNF. Then regions of constant offset start to grow from the boundaries towards the interior of the inpainting domain, creating a (rather simple) partition of $\widetilde{O}$.

Let us analyze how would the inpainting look like for the simple case in which O is partitioned in two regions of constant ONNF, R_1 with $t(x) = t_1$ and R_2 with t_2

Figure 8.5: **Results on periodic textures.** KT: method of Komodakis and Tziritas [39] and KSY: method of Kawai *et al.* [37]. M and P stand for patch NL-means and -Poisson. From top to bottom *matsuri* and *mailboxes*.

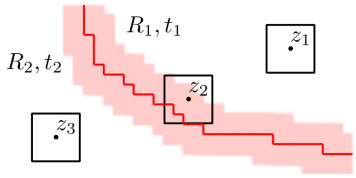

Figure 8.6: **Two copy regions and the transition band.** In regions R_1 and R_2, which have a constant NNF, data is rigidly translated (copied) from corresponding source regions in the complement. The transition between these copy regions takes place on a band whose width coincides with the patch size.

(Figure 8.6). Thus, for the patch NL-means we have

$$u(z) = \alpha_1 \hat{u}(z + t_1) + \alpha_2 \hat{u}(z + t_2),$$

where $\alpha_i = \int_{\Omega_p} g_a(y) \chi_{R_i}(z - y) \mathrm{d}y$, $i = 1, 2$, is the area of the intersection of the patch centered at z with R_i.

Pixels in the red band in Figure 8.6 receive two contributions ($\alpha_1, \alpha_2 > 0$). Outside this band, in both regions R_1 and R_2, the image u results from a rigid translation (*i.e.* a *verbatim* copy) from two corresponding regions in O^c. The transition between both *copy regions* takes place at the red *transition band*. Patch NL-means performs a smooth blending.

This argument generalizes to an arbitrary number of regions. The value of u at each pixel z is determined by the copy regions overlapped by the patch centered at z, weighted by the overlap area. The transition bands are defined as the centers of patches intersecting at least two different copy regions. Outside these bands, the resulting image is an exact copy (of intensities or gradients) of a corresponding source region in O^c.

The bottom row in Figure 8.7 shows the evolution of the patch error $\varepsilon(p_u(x) - p_{\hat{u}}(n(x)))$. Recall from (2.9) that, in the limit $T \to 0$, the energy is computed as the sum of these errors. The energy is concentrated around the transition bands, since patches that do not overlap any band are an exact copy of the source patch.

This explains why the minimization of the energy often favors the emergence of copy regions. Let us point out that there are cases in which the optimum inpainting does not present any copy region at all. The multiscale approach tends to prevent convergence to these kind of minima.

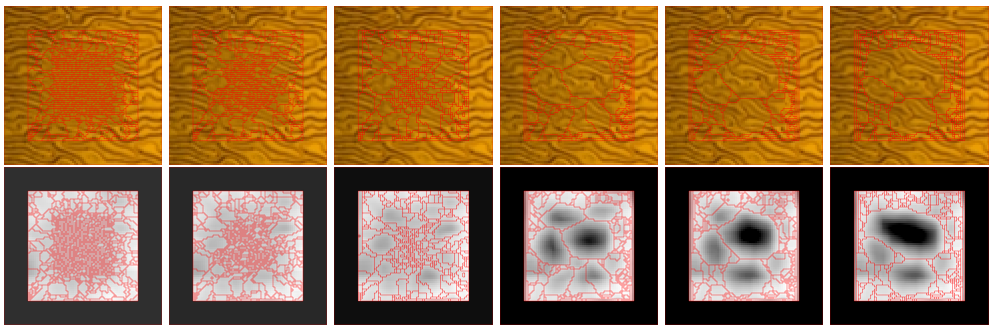

Figure 8.7: **Minimization of the energy.** Iterations $k = 1, 2, 3, 20, 50, 66$ of the minimization procedure corresponding to the patch NL-means (all schemes have a similar behaviour). The algorithm converged at $k = 66$. The bottom row shows the corresponding distribution of the patch error (energy density). Notice the emergence of coherent copy regions and how the energy concentrates along their boundaries.

8.3. Some comments. The patch NL-Poisson performs well in images with a strong structure (Figures 8.3 and 8.5) but fails in images characterized by random textures. In the following we are going to analyze the reasons for this behavior, discussing the benefits and limitations of using gradients both in the image and weights update steps.

Gradients in the image synthesis. The image obtained with patch NL-Poisson is the result of copying gradients from the known portion of the image into the hole and then solving a PDE. Although we do not have a proof of it, in our experiments we observe that the synthesized image will not have edges which are not in O^c. This is not the case for patch NL-means, which may present discontinuities (seams) at the boundary of the hole (see for instance the result of patch NL-means in *sofa*, Figure 8.3).

Gradients in the patch similarity weights. For patch NL-Poisson (8.1), the patch error is a combination of intensity and gradients. With the low values of λ used, the gradient component dominates. For some textured images this may cause the method to fail. For instance, in *baseball* (Figure 8.4), segments of the sky have been reproduced in the snow. The result with $\lambda \approx 1$ produces a better reconstruction.

8.4. Further developments. Let us discuss some preliminary results on several directions that deserve future exploration.

Equilibria of some non-variational schemes. Figure 8.8 shows results obtained with a variant of the patch NL-Poisson. Different values of the mixture coefficient λ are used for the image (λ_u) and weights update (λ_w). Results in Figure 8.8 correspond to $\lambda_w = 1$, *i.e.* the weights are computed based only on the image values. The image is updated using the corresponding image update step with a low value of λ_u. Such scheme is non-variational unless $\lambda_u = \lambda_w$. However, it can be proven that the iterative scheme converges to a Nash (type) equilibrium of two different energy functionals, one

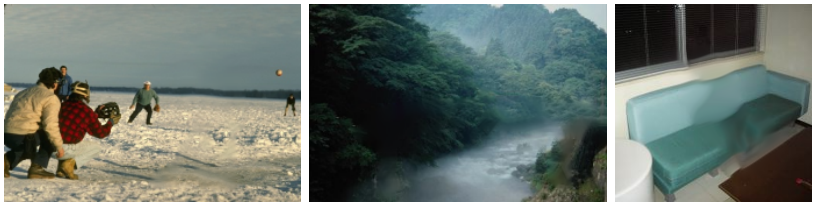

Figure 8.8: **Results for non-variational gradient methods**. Results using the non variational extension of the patch NL-Poisson. The patch similarity is computed based solely on image values (setting $\lambda_w = 1$), whereas the image update step is mainly gradient-based by keeping λ_u to a low value.

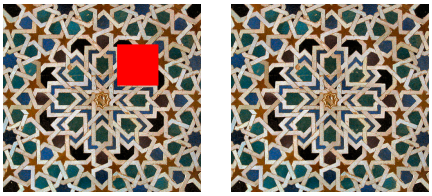

Figure 8.9: **Symmetries**. Result using patch NL-means. To enrich the set of exemplars, $\widetilde{O}^c$ is redefined as the union of the complement of the inpainting domain with its vertical, horizontal and central symmetries.

for the computation of weights when the image is fixed, and the other for the synthesis of the image, when the weights are fixed.

Enriching the set of exemplars. Our inpainting formalism can be easily extended to handle transformations of exemplars (such as symmetries, rotations, or affinities), by considering an orbit of $\widetilde{O}^c$:

$$\mathcal{E}(u,w) = \sum_{\tau\in\mathcal{T}} \int_{\widetilde{O}} \int_{\tau^{-1}(\widetilde{O}^c)} w(x,y,\tau)\varepsilon(p_u(x) - p_{\hat{u}\circ\tau}(y))\mathrm{d}y\mathrm{d}x - T\int_{\widetilde{O}} \mathcal{H}(w(x,\cdot,\cdot))\mathrm{d}x$$

where $\mathcal{T}$ is the family of transformation defining the orbit. Here $w(x,\cdot,\cdot)$ is a probability over the orbit. In Figure 8.9 we show a preliminary result considering as the orbit, the identity and vertical, horizontal and central symmetries. From a computational point of view, this is equivalent to having an image with a larger $\widetilde{O}^c$. It is interesting to point out, that since the computational cost of the PatchMatch is of $O(|\Omega_p||\widetilde{O}|)$, this has little effect on the computing time.

Deterministic Annealing. Our formalism is related to the deterministic annealing framework for clustering of [51]. In this work the author presents a *deterministic annealing* scheme for finding a global minimum for an energy closely related to ours. In our context, this corresponds to minimization of a series of energies $(\mathcal{E}_{2,T_n})_n$ where T_n is a decreasing sequence of temperatures (*e.g.* $T_n = \alpha^n T_0$). The minimum find for T_n is used as initialization for T_{n+1}.

Our context is not equivalent to that of [51], however it is still interesting to explore the application of such an annealing scheme. Indeed, this has already been used for non-local demosaicing [1] and in our previous work on interpolation of sparsely sampled images [28].

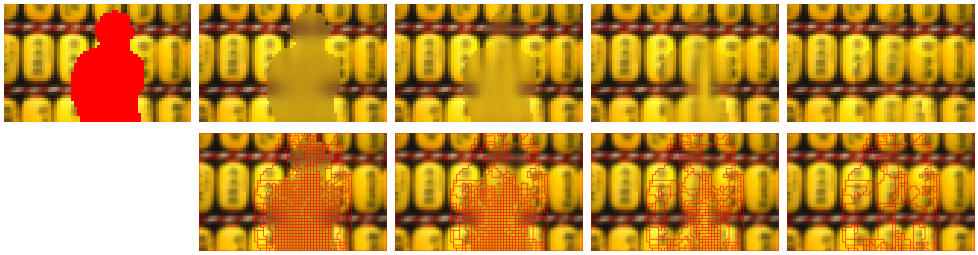

Figure 8.10: **Deterministic annealing**. Result using patch NL-means. Top: original image and inpainting domain. Some iterates of a deterministic annealing process, from left to right $T = 500, 154, 65, 61, 1$.

Figure 8.10 shows some images of the annealing sequence for $T = 500, 154, 65, 61$. The red curves on the bottom row depict the boundaries of the regions with constant offset with respect to the nearest neighbor. This offset is given by $t(x) = n(x) - x$, for $x \in \widetilde{O}$, where $n(x) \in \widetilde{O}^c$ denotes the position of the nearest neighbor of $p_u(x)$. When $T = 0$, $n(x)$ corresponds to the correspondence map φ. However, it is still useful to show the properties of the solution with $T > 0$.

When decreasing T, the results vary smoothly, except at some critical temperatures in which the solution changes considerably, suggesting the possibility of phase transitions. This can be seen in Figure 8.10: $T = 65$ and $T = 61$ correspond to consecutive temperatures in the sequence. Their solutions differ considerably. Notice also the corresponding change in the regions of constant offset t.

The results seem to support the idea that the optimal correspondence maps obtained by annealing have the regularity properties described in Theorem 5.6. We do not know the answer to this question.

9. Appendix: proof of Lemma 5.4. Assuming that Ω_p is an open set, then $\widetilde{O}$ is also open, hence $\widetilde{O}^c$ is closed. Let us denote by $M(\widetilde{O}^c)$ the Banach space of (finite) Radon measures in $\widetilde{O}^c$ which is the dual of the space of continuous functions in $\widetilde{O}^c$, denoted by $C(\widetilde{O}^c)$. We denote by $\sigma(M(\widetilde{O}^c), C(\widetilde{O}^c))$ the weak* topology of $M(\widetilde{O}^c)$.

Proof. Clearly, the measures ν^φ are extreme points of $\mathcal{MP}$. Let $\nu \in \mathcal{MP}$ be an extreme point. Then for almost any $x \in \widetilde{O}$, ν_x is also an extreme point of the set of probability measures in $\widetilde{O}^c$. Let us prove this assertion. The decomposition of a measure λ into its diffuse part λ^d and its atomic part λ^a as well as the Lebesgue decomposition $\lambda = \lambda^{ac} + \lambda^s$ into its $\mathcal{L}^N$-absolutely continuous part λ^{ac} and its λ^s-singular part define maps $\lambda \to \lambda^d$, $\lambda \to \lambda^a$, $\lambda \to \lambda^{ac}$ and $\lambda \to \lambda^s$ in $M(\widetilde{O}^c)$, which are measurable with respect to the Borel structure generated by the $\sigma(M(\widetilde{O}^c), C(\widetilde{O}^c))$-topology on $M(\widetilde{O}^c)$ [24] (see also [56]). Thus the maps $x \to \nu_x^a, \nu_x^{ac}, \nu_x^s$ are measurable.

Let us observe that we can reduce the proof to the case that only one of the components is non-null. Indeed, let us assume that we can write ν_x as

$$\nu_x = \alpha(x)\nu_x^A + \beta(x)\nu_x^B,$$

where $\nu_x^A, \nu_x^B \in \mathcal{MP}$, ν_x^A and ν_x^B are orthogonal, and $\alpha(x), \beta(x) \geq \alpha > 0$ on some set $X \subseteq \widetilde{O}$ of positive (Lebesgue) measure. Then we define

$$\nu_x^1 = (\alpha(x) - \alpha\chi_X(x))\nu_x^A + (\beta(x) + \alpha\chi_X(x))\nu_x^B,$$

$$\nu_x^2 = (\alpha(x) + \alpha\chi_X(x))\nu_x^A + (\beta(x) - \alpha\chi_X(x))\nu_x^B,$$

Then $\nu_x^1, \nu_x^2 \in \mathcal{MP}$ and $\nu_x = \frac{1}{2}\nu_x^1 + \frac{1}{2}\nu_x^2$. Since ν_x^A and ν_x^B are orthogonal this implies that ν is not an extreme point of $\mathcal{MP}$. Thus we may assume that $\alpha \wedge \beta = 0$ (the operator $\wedge$ acting on two measures denotes its infimum as measures [3]) and we may consider that ν_x has only one component.

Let us consider the case where $\nu_x = \nu_x^s$ a.e. in x. Let us prove that we can find sets $A, B \subseteq \widetilde{O}^c$ such that

$$\{x \in \widetilde{O} : \nu_x(A) > 0, \nu_x(B) > 0\} \text{ is of positive Lebesgue measure.} \tag{9.1}$$

Recall that since ν_x are singular measures $\nu_x(B) \to 0$ as diameter$(B) \to 0+$. Let us consider a dyadic subdivision of the minimum rectangle containing $\widetilde{O}^c$. We consider the rectangles of the subdivision so that they are of the form $[a, b) \times [c, d)$. As the subdivision gets refined their diameter tends to $0+$. At the first step $i = 0$ of the subdivision we have four sets: $D_i^1, D_i^2, D_i^3, D_i^4$. Thus either our assertion is true or we may divide $\widetilde{O}$ into four disjoint sets $E_i^r = \{x \in \widetilde{O} : \text{the support of } \nu_x \text{ is contained in } D_i^r\}$, $r = 1, 2, 3, 4$. Thus if our assertion was never true, we would get that the support of each ν_x is contained in a set whose diameter tends to 0, therefore $\nu_x = 0$ a.e. in x. Thus (9.1) is true. Thus we may assume that

$$\nu_x = \alpha(x)\nu_x^A + \beta(x)\nu_x^B,$$

where $\nu_x^A = \frac{\chi_A\nu_x}{\|\chi_A\nu_x\|}$, $\nu_x^B = \frac{\chi_B\nu_x}{\|\chi_B\nu_x\|}$, $\alpha(x) = \|\chi_A\nu_x\|$, $\beta(x) = \|\chi_B\nu_x\|$. Proceeding as in the previous proof where we reduced to the case of only one component we prove that ν cannot be an extreme point of $\mathcal{MP}$.

In a similar way, we prove that if $\nu_x = \nu_x^{ac}$ a.e. in x, then ν cannot be an extreme point of $\mathcal{MP}$. But, let us notice that in the present case we can give a more simple proof. Let us prove that $\nu_x^{ac} = 0$ a.e.. Otherwise, let $\alpha_x > 0$ be such that $\int_{\widetilde{O}^c}(\nu_x^{ac} - \alpha_x)\,dy = \frac{1}{2}\int_{\widetilde{O}^c}\nu_x^{ac}\,dy$. Then also $\int_{\widetilde{O}^c}(\nu_x^{ac} \wedge \alpha_x)^+\,dy = \frac{1}{2}\int_{\widetilde{O}^c}\nu_x^{ac}\,dy$. Then

$$\nu_x^{ac} = \frac{1}{2}(2(\nu_x^{ac} - \alpha_x)^+) + \frac{1}{2}(2(\nu_x^{ac} \wedge \alpha_x)),$$

and ν cannot be an extreme point of $\mathcal{MP}$.

We conclude that $\nu_x = \nu_x^a$ a.e.. Assume that $\nu_x^a = \alpha_1(x)\delta_{\sigma_1(x)} + \alpha_2(x)\delta_{\sigma_2(x)} + \nu_x'$ [36, 56]. We proceed as above: We may assume that there is $\alpha > 0$ such that $\alpha_1(x), \alpha_2(x) \geq \alpha$ on some measurable set $A \subseteq \widetilde{O}$. Then we define $\nu_x^{a1} = (\alpha_1(x) - \alpha\chi_A(x))\delta_{\sigma_1(x)} + (\alpha_2(x) + \alpha\chi_A(x))\delta_{\sigma_2(x)} + \nu_x'$, $\nu_x^{a2} = (\alpha_1(x) + \alpha\chi_A(x))\delta_{\sigma_1(x)} + (\alpha_2(x) - \alpha\chi_A(x))\delta_{\sigma_2(x)} + \nu_x'$. Then $\nu_x^a = \frac{1}{2}\nu_x^{a1} + \frac{1}{2}\nu_x^{a2}$. Thus ν_x^a cannot be an extreme point unless there exists a map $x \in \widetilde{O} \to \varphi(x) \in \widetilde{O}^c$ such that $\nu_x = \delta_{\varphi(x)}(y)$. By the observation previous to the Lemma, φ is measurable. Hence $\nu = \nu^\varphi$. □

Acknowledgements. We acknowledge partial support by MICINN project, reference MTM2009-08171, and by GRC reference 2009 SGR 773. PA is supported by the FPI grant BES-2007-14451 from the Spanish MICINN. VC also acknowledges partial support by IP project "2020 3D Media: Spacial Sound and Vision", financed by EC, and by "ICREA Acadèmia" prize for excellence in research funded both by the Generalitat de Catalunya.

REFERENCES

[1] Buades A., Coll B., Morel J.-M., and Sbert C. Self-similarity driven color demosaicking. *IEEE Trans. on IP*, 18(6):1192–1202, 2009.
[2] C.D. Aliprantis and K.C. Border. *Infinite dimensional analysis: a hitchhiker's guide*. Springer Verlag, 2006.
[3] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford University Press, USA, 2000.
[4] P. Arias, V. Caselles, and G. Sapiro. A variational framework for non-local image inpainting. In *EMMCVPR*, Lecture Notes in Computer Science, pages 345–58. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
[5] P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93:319–347, 2011.
[6] J.-F. Aujol, S. Ladjal, and S. Masnou. Exemplar-based inpainting from a variational point of view. *SIAM J. Math. Anal.*, 42(3):1246–85, 2010.
[7] S.P. Awate and R.T. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. on PAMI*, 28(3):364–376, 2006.
[8] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans. on IP*, 10(8):1200–11, 2001.
[9] C. Barnes, E. Shechtman, A. Finkelstein, and D. B Goldman. PatchMatch: a randomized correspondence algorithm for structural image editing. In *Proc. of SIGGRAPH*, pages 1–11, New York, NY, USA, 2009. ACM.
[10] C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized PatchMatch correspondence algorithm. In *European Conference on Computer Vision*, 2010.
[11] M. Bertalmío, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proc. of SIGGRAPH*, 2000.
[12] M. Bertalmío, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture inpainting. *IEEE Trans. on Image Processing*, 12(8):882–89, 2003.
[13] D.P. Bertsekas. *Nonlinear programming*. Athena Scientific Belmont, MA, 1999.
[14] R. Bornard, E. Lecan, L. Laborelli, and J.-H. Chenot. Missing data correction in still images and image sequences. In *Proc. ACM Int. Conf. on Multimedia*, 2002.
[15] F. Bornemann and T. März. Fast image inpainting based on coherence transport. *Journal of Mathematical Imaging and Vision*, 28(3):259–78, 2007.
[16] T. Brox, O. Kleinschmidt, and D. Cremers. Efficient nonlocal means for denoising of textural patterns. *IEEE Trans. on IP*, 17(7):1057–92, 2008.
[17] A. Buades, B. Coll, and J.M. Morel. A non local algorithm for image denoising. In *Proc. of the IEEE Conf. on CVPR*, volume 2, pages 60–65, 2005.
[18] F. Cao, Y. Gousseau, S. Masnou, and P. Pérez. Geometrically guided exemplar-based inpainting. 2009.
[19] T. Chan, S. H. Kang, and J. H. Shen. Euler's elastica and curvature based inpaintings. *SIAM J. App. Math.*, 63(2):564–92, 2002.
[20] T. Chan and J. H. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. App. Math.*, 62(3):1019–43, 2001.
[21] A. Criminisi, P. Perez, and K. Toyama. Region filling and object removal by exemplar-based inpainting. *IEEE Trans. on Image Processing*, 13(9):1200–1212, 2004.
[22] L. Demanet, B. Song, and T. Chan. Image inpainting by correspondence maps: a deterministic approach. *Applied and Computational Mathematics*, 1100:217–50, 2003.
[23] I. Drori, D. Cohen-Or, and H. Yeshurun. Fragment-based image completion. *ACM Trans. on Graphics. Special issue: Proc. of ACM SIGGRAPH*, 22(3):303–12, 2003.
[24] L. Dubins and D. Freedman. Measurable sets of measures. *Pacific Journal of Mathematics*, 14(4):1211–1222, 1964.
[25] A.A. Efros and T.K. Leung. Texture synthesis by non-parametric sampling. In *Proc. of the IEEE Conf. on CVPR*, pages 1033–1038, Corfu, Greece, September 1999.
[26] R.S. Ellis. *Entropy, large deviations, and statistical mechanics*, volume 1431. Springer Verlag, 2005.
[27] S. Esedoglu and J. H. Shen. Digital image inpainting by the mumford-shah-euler image model. *European J. App. Math.*, 13:353–70, 2002.
[28] G. Facciolo, P. Arias, V. Caselles, and G. Sapiro. Exemplar-based interpolation of sparsely sampled images. In *EMMCVPR*, Lecture Notes in Computer Science, pages 331–44. Springer Berlin Heidelberg, 2009.
[29] D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. Springer Verlag, 2001.
[30] G. Gilboa and S. Osher. Nonlocal operators with applications to image processing. *Multiscale Model. Simul*, 7(3):1005–1028, 2008.

[31] G. Gilboa and S.J. Osher. Nonlocal linear image regularization and supervised segmentation. *SIAM Mult. Mod. and Sim.*, 6(2):595–630, 2007.
[32] P. Harrison. *Texture tools.* PhD thesis, Monash University, 2005.
[33] H. Igehy and L. Pereira. Image replacement through texture synthesis. In *Proc. of the IEEE Conf. on CVPR*, October 1997.
[34] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–30, 1957.
[35] J. Jia and C.-K. Tang. Inference of segmented color and texture description by tensor voting. *IEEE Trans. on PAMI*, 26(6):771–86, 2004.
[36] NJ Kalton. The endomorphisms of Lp, 0? p? 1. *Indiana Univ. Math. J*, 27(3):353–381, 1978.
[37] N. Kawai, T. Sato, and N. Yokoya. Image inpainting considering brightness change and spatial locality of textures and its evaluation. In *Ad. in Image and Video Tech.*, pages 271–82. Springer Berlin Heidelberg, 2009.
[38] S. Kindermann, S. Osher, and P.W. Jones. Deblurring and denoising of images by nonlocal functionals. *Multiscale Modeling and Simulation*, 4(4):1091–1115, 2006.
[39] N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Trans. on IP*, 16(11):2649–61, 2007.
[40] V. Kwatra, I. Essa, A. Bobick, and N. Kwatra. Texture optimization for example-based synthesis. *ACM Trans. on Graph.*, 24(3):795–802, 2005.
[41] A. Levin, A. Zomet, and Y. Weiss. Learning how to inpaint from global image statistics. In *Proc. of IEEE ICCV*, 2003.
[42] E. Levina and P. Bickel. Texture synthesis and non-parametric resampling of random fields. *Annals of Statistics*, 34(4), 2006.
[43] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Trans. on IP*, 11(2):68–76, 2002.
[44] S. Masnou and J.-M. Morel. Level lines based disocclusion. In *Proc. of IEEE ICIP*, 1998.
[45] P. Perez, M. Gangnet, and A. Blake. PatchWorks: Example-based region tiling for image editing. Technical report, Microsoft Research, 2004.
[46] G. Peyré. Manifold models for signals and images. *To appear in Comp. Vis. and Im. Unders.*
[47] G. Peyré, S. Bougleux, and L. Cohen. Non-local regularization of inverse problems. In *ECCV '08*, pages 57–68, Berlin, Heidelberg, 2008. Springer-Verlag.
[48] G. Peyré, S. Bougleux, and L. D. Cohen. Non-local regularization of inverse problems. *Preprint Hal-00419791*, 2009.
[49] L. Pizarro, P. Mrázek, S. Didas, S. Grewenig, and J. Weickert. Generalised nonlocal image smoothing. *International Journal of Computer Vision*, 90(1):62–87, 2010.
[50] M. Protter, M. Elad, H. Takeda, and P. Milanfar. Generalizing the non-local-means to super-resolution reconstruction. *IEEE Trans. on IP*, 18(1):36–51, 2009.
[51] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 86(11):2210–2239, 1998.
[52] S.M. Srivastava. *A course on Borel sets*, volume 180. Graduate Texts in Mathematics, Springer Verlag, 1998.
[53] J. Sun, L. Yuan, J. Jia, and H. Y. Shum. Image completion with structure propagation. In *Proc. of SIGGRAPH*, New York, NY, USA, 2005. ACM.
[54] D. Tschumperlé and R. Deriche. Vector-valued image regularization with pde's: a common framework for different applications. *IEEE Trans. on PAMI*, 27(4), 2005.
[55] L.Y. Wei and M. Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proc. of the SIGGRAPH*, 2000.
[56] L. Weis. On the Representation of Order Continuous Operators by Random Measures. *Transactions of the American Mathematical Society*, 285(2):535–563, 1984.
[57] Y. Wexler, E. Shechtman, and M. Irani. Space-time completion of video. *IEEE Trans. on PAMI*, 29(3):463–476, 2007.
[58] C.F.J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, pages 95–103, 1983.