ANALYSIS OF THE WAVELET GINZBURG–LANDAU ENERGY IN IMAGE APPLICATIONS WITH EDGES

J.A. DOBROSOTSKAYA AND A.L. BERTOZZI

Abstract. A wavelet analogue of the Ginzburg–Landau energy (WGL) was recently designed and integrated in variational methods for image processing. In this paper we prove global well–posedness of the gradient descent equation (in the weak sense) and convergence to an extremum. We also develop further uses for this energy completed with an additional edge preserving forcing term. We present examples including inpainting, superresolution, segmentation, denoising, and contour detection. Combining the spatial and the edge-preserving forcing terms gives an amazing flexibility of the model, that is not only applicable to various image processing tasks, but also highly tunable.

1. Introduction. This paper discusses variational properties of the wavelet Ginzburg–Landau energy ([11, 12]) and well-posedness of the associated gradient descent problem. We investigate the properties of the minimizers in the diffuse-interface context. We also continue our study of the qualitatively new category of methods that use wavelets in settings analogous to the widely known differential models for image processing. Its key elements are the nonlocal operators based on wavelets (as a particular case of a sparse representation system). One can see it as a way to combine advantages of two effective and popular classes of techniques within one method, which may also lead to a wider acceptance of the variational methods by considering those not involving numerical PDE solving.

The wavelet Ginzburg-Landau energy was first introduced in an earlier work [11] that described the design of the wavelet Laplacian — a pseudo–differential operator that retains the properties of its differential prototype (and brings in additional advantages) in the context of an image processing model involving diffusive structure. This idea was inspired by [4], where the properties of Cahn–Hilliard equation were utilized for image inpainting, more specifically — for restoration of broken connections over the missing image parts. The wavelet Laplacian actually inherits many properties of the regular Laplace operator, moreover, just as the associated WGL functional matches the role of the classical Ginzburg-Landau energy(GL), while bringing in significant advantages [11] such as introducing much lower diffuse-interface blur in the processed images, nevertheless restoring the feature connections. In another work [12] we proved that in the case of 2D binary functions the wavelet Ginzburg Landau energy (defined using a regular wavelet) converges in the variational sense (Γ-converges) to its sharp-interface limit as the analogue of the interface parameter ϵ tends to 0. However, unlike the Total Variation functional, that is the sharp limit of the classical Ginzburg–Landau energy, the limiting energy for WGL is anisotropic, the anisotropy arising from the structure of a separable dyadic wavelet kernel used in the definition of the WGL energy.

Although the rigorous analysis of the Γ-convergence had been done for this energy in [12], in practice we use gradient descent to compute the energy minimizers, which makes it necessary to develop theory for the associated dynamic problem. This is the first paper that addresses this issue and discusses the properties of the respective solutions in the diffuse-interface context. After we justify the well-posedness of the problem and properties of the resulting minimizers, we apply the gradient descent method in numerical simulations of image recovery.

This paper describes a generalization of the inpainting model from [11] that becomes a more unified image recovery tool. It is also variational, and uses the WGL

energy as the regularizing part, but includes two forcing terms in the energy that is subject to minimization: the edge preserving fidelity term and the spatial fidelity term that already appeared in the old model. This additional flexibility in incorporating any given information in the problem setting extends the applicability of this model to many more image reconstruction applications than the classically defined image inpainting. Namely, this model can be used for the recovery of noisy, occluded or incompletely transmitted binary images, for segmentation of images (including binary contour extraction), for inpainting, superresolution and other types of grayscale and color image processing. Examples of multipurpose image reconstruction/analysis are addressed from a unified point of view originating from the WGL image inpainting technique.

The problem of image reconstruction from incomplete data has been studied extensively from various angles. Two important classes of relevant techniques are those addressing the recovery of pixel information and wavelet coefficients. Restoration of missing or occluded parts in the spatial (pixel) domain, generally referred to as "image inpainting", is a problem of a related class, formulated and studied mostly during the last decade (see [2],[1],[3],[6] and other works). The term "inpainting" is self–explanatory: indeed, numerical techniques suitable for this problem tend to inherit visual properties of hand–drawn painting restorations. A separate class of image recovery methods are designed to manipulate the pixel values implicitly, via processing performed entirely in the wavelet domain ([8],[7],[15]). Both of the mentioned categories of methods often involve PDE–based processing in either the pixel or wavelet domain. The variational method based on the WGL energy is qualitatively different: rather than expressing the discretized versions of the differential operators using wavelets([21]), or applying the known variational methods in the wavelet domain([8] energy and associated "diffusion-style" operators that replace and outperform the differential operators in a diffuse-interface setting.

The paper is organized as follows. After introducing main notions, we discuss some properties of the wavelet Laplace operator (Section 2.1). Then we address the gradient descent minimization of the WGL energy from the "classical PDE" point of view and prove that the respective differential equation is well–posed and has global in time solutions converging to extrema of WGL (Section 2.2). We also give some a priori regularity estimates of the minimizers as well as some sufficient conditions for the minimizer uniqueness. The rest of the paper is devoted to the discussion of variational techniques for multipurpose image processing based on the minimization of WGL with additional forcing terms. In this context we introduce the edge–preserving fidelity term (section 3.1) and discuss its construction depending on the image processing application involved. The presence of both the spatial and the edge-preserving wavelet forcing terms enhances the model by making it adaptable to a variety of image recovery tasks and tunable within each of those.

2. Variational properties of the WGL functional.

2.1. Preliminary discussion of the problem setting. Let us introduce the main notations and assumptions that will be used throughout the text.

We consider functions defined periodically on $[0,1]^N$ (unless otherwise specified). Function spaces are defined accordingly (e.g. L^2 means $L^2([0,1]^N)$).

The sign of the scalar product $\langle \cdot, \cdot \rangle$ denotes the L^2 product unless otherwise specified.

Whenever the limits of integration are omitted, we assume the integration over the entire domain.

Consider the class of wavelets ψ that have an associated scaling function ϕ. Define the wavelet mode (j, k) as

$$\psi_{j,k}(x) = 2^{Nj/2}\psi(2^j x - k), \quad j = 0, 1, 2, \ldots, k \in \mathbb{R}^N,$$

and the wavelet transform of a function $f \in L^2$ at the mode (j, k) as

$$Wf(j, k) = \langle f, \psi_{j,k} \rangle. \tag{WT}$$

Similarly, let us also denote

$$\phi_{j,k}(x) = 2^{Nj/2}\phi(2^j x - k), \quad j = 0, 1, 2, \ldots, k \in \mathbb{R}^N.$$

Later in the text we will make additional assumptions on the regularity of the wavelet ψ. In the case of functions defined on $[0, 1]^N$ the range of the translation parameter k at the $j-th$ dilation scale is restricted to $[0, 2^j]^N$. When discussing multi-dimensional cases, we use ψ as a general notation for the wavelet functions, assuming, wherever needed, summation over all of those.

Image processing techniques, in most cases, need to be translation invariant. We will be using the *semicontinuous dyadic wavelet transform* ([17]), which produces the following decomposition for any function $u \in L^2$:

$$u(x) = \int \langle u, \phi_{0,\mu} \rangle \phi_{0,\mu} d\mu + \sum_j \int \langle u, \psi_{j,\mu} \rangle \psi_{j,\mu} d\mu. \tag{IWT}$$

We will denote an orthogonal projection of a function u on the subspace Λ as $Pr_\Lambda u$, and also use the following notation for the continuous analogues of projections of an arbitrary function $u \in L^2$ on the wavelet–generated "approximation" or "detail" subspaces V_j or W_j (as defined in [20]):

$$P_{W_j}u = \int \langle u, \psi_{j,\mu} \rangle \psi_{j,\mu} d\mu, \quad P_{V_j}u = \int \langle u, \phi_{0,\mu} \rangle \phi_{0,\mu} d\mu + \sum_{s=0}^{j-1} P_{W_s}u.$$

We will also use $P_\Lambda u$ to denote an analogous operator for an arbitrary subspace Λ.

When we consider periodic functions on a finite rectangular domain, without further adjustment of notations, we assume that the wavelet transform is also periodized (details on the wavelet kernel periodization can be found in [17], [18] or [20]). Notice that in the case of the non-redundant discrete representation, we will use the following projection notations:

For any function $u \in L^2$ we define its B-seminorm in the translation invariant form as

$$|u|_B = \left(\sum_{j=0}^{\infty} 2^{2j} \int |\langle \psi_{j,\mu}, u \rangle|^2 d\mu \right)^{1/2}.$$

This seminorm is equivalent to the classical, translation-dependent Besov $(B_{2,2}^1)$ seminorm on the cases when the wavelet ψ is sufficiently regular ([11]). In the latter case the translation-invariant version of the Besov seminorm equal to the average of the classical Besov seminorms of the given function shifted over the unit interval. We also extend this definition and consider *Besov–type seminorms*, which are defined as above, except for the wavelet ψ being non-regular.

In other words, we define a Banach space B as a space of Lebesgue measurable functions with a norm expressed directly in terms of the wavelet coefficients:

$$\|u\|_B^2 = \|P_{V_0} u\|_{L^2}^2 + |u|_B^2.$$

If ψ is a sufficiently regular wavelet(r-regular [18] with $r \geq 2$), B is the Besov 1-2-2 space, which is equivalent to the Sobolev space H^1 with the equivalence of the respective seminorms:

$$|u|_B \asymp |u|_{H^1},$$

where

$$|u|_{H^1}^2 = \int |\nabla u(x)|^2 dx.$$

We will simplify notations further by calling such wavelets "regular". For other wavelets, which we will call "non–regular", this definition produces a "weighted L^2" Besov–type space, that extends the definition of Besov spaces formally: the norm and seminorm are defined in terms of the wavelet coefficients as described above. In fact, every space B generated by any non-regular wavelet consists of piece-wise constant functions. Those are identically constant on their domain in case one uses the translation invariant B-seminorm, and include functions with jumps on the dyadic grid in case the seminorm is defined using the non-redundant wavelet transform.

Our interest in the Besov–type spaces is related to the applications that require the wavelet generating the WGL energy to have a special shape. An example of such application is the blind deconvolution of the 2D bar codes (see [10]). Therefore, when talking about the Besov–type spaces one can always replace those with finite–scale subspaces of those that are encountered in practice when processing finite signals (images).

To explain the nature of our computational examples as well as to illustrate the case of the WGL energy on the Besov–type spaces, we will also consider the following finite–dimensional (hence, associated with a discrete representation) spaces

$$\mathcal{V}_J = \{u(x) = \sum_{\vec{k}:k_i=0,1,\dots 2^J-1} \{a_{\vec{k}} \phi_{J,\vec{k}}\}\}.$$

For example, if ψ is the Haar wavelet function, then $\mathcal{V}_J$ consists of piecewise constant functions discontinuous only on the dyadic grid of mesh size 2^{-J}. The space $\mathcal{V}_J$ of functions of two variables generated by any separable wavelet kernel has periodic dyadic structure suitable for image processing. The following representation of any function $u \in \mathcal{V}_J$ is translation invariant in the discrete sense, i.e. invariant w.r.t. translations by multiples of 2^{-J}:

$$u(x) = \sum_{\vec{k}:k_i=0,1,\dots 2^J-1} \langle \psi_{J,\vec{k}}, u \rangle \psi_{J,\vec{k}}(x). \tag{U_d}$$

This representation can be viewed as an integral representation w.r.t. the singular measure
$\sum_{\vec{k}:k_i=0,\dots,2^J-1} \delta(\vec{k}2^{-J})$. Respectively, the B seminorm can be re-defined for functions

from such spaces in a discrete (but still redundant) form as

$$|u|_{B(\mathcal{V}_J)} = \left(\sum_{j=0}^{J} 2^{2j} \sum_{\vec{k}:k_i=0,\ldots,2^J-1} |\langle u, \psi_{j,k}\rangle|^2 d\mu \right)^{1/2}. \qquad (B_d)$$

So, given an image f of size $2^N \times 2^N$ we assume that the intensity values $f_{i,j}$ at the pixel position (i,j) equal to the respective coefficients $a_{\vec{k}}$, $\vec{k} = (i,j)^T$ up to a normalizing (rescaling) constant. All of our numerical examples were implemented in Matlab and use the discrete representation (U_d) with the wavelet coefficient values obtained from the stationary wavelet transform (swt) routine.

For the purposes of our discussion, we will always assume that we operate with an orthonormal wavelet ψ that has a scaling function ϕ and assume that the associated B-seminorm is translation invariant (in the discrete or continuous sense). To simplify notations, we will not distinguish between those, since all theoretical arguments stay valid for both cases (unless otherwise specified).

WAVELET GINZBURG–LANDAU ENERGY. The main object of analysis as well as the key tool for image processing applications considered in this paper is the wavelet Ginzburg–Landau (WGL) energy functional (introduced in [11])

$$WGL(u) = \frac{\epsilon}{2}|u|_B^2 + \frac{1}{4\epsilon} \int W(u)dx, \quad \text{where } W(u) = (u^2 - 1)^2.$$

We will use an operator named the "wavelet Laplacian" (introduced in [11] in terms of the discrete wavelet transform), which has the wavelet basis functions as eigenfunctions, and acts on those in the same "scale–proportional" manner as the Laplace operator does on the Fourier basis. Given an orthonormal wavelet ψ the "wavelet Laplacian" of any $u \in L^2(\mathbb{R})$ is formally defined as

$$\Delta_w u = -\sum_{j=0}^{+\infty} 2^{2j} \int \langle f, \psi_{j,\kappa}\rangle \psi_{j,\kappa} d\kappa. \qquad (WvL)$$

In the case of $\mathcal{V}_J$, we also define

$$\Delta_w^{\mathcal{V}_J} u = -\sum_{j=0}^{+\infty} 2^{2j} \sum \langle f, \psi_{j,\kappa}\rangle \psi_{j,\kappa} d\kappa.$$

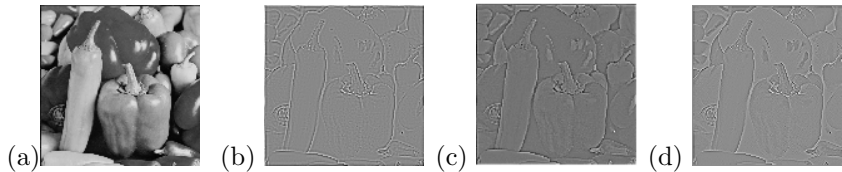Fig.2.1 compares the regular and wavelet Laplace operators applied to the "peppers" image.

(a)   (b)   (c)   (d)

FIG. 2.1. *(A) Original "peppers" image, (B) wavelet Laplace operator generated by DB4 wavelet, (C) wavelet Laplace operator generated by Haar wavelet, (D) regular Laplace operator.*

Translation invariant energy, as well as an operator based on the redundant wavelet representation, cause certain difficulty in the problem analysis, since the semi–discrete dyadic wavelet transform, even though associated with a stable frame, is highly redundant. There is no orthogonality between the terms of the function representation. Remarkably, the derivative-style identity $\langle -\Delta_w u, u \rangle = |u|_B^2$ still holds. Indeed, in the case of functions defined on $\mathbb{R}^N$, rewriting the product in terms of the Fourier transforms gives

$$\langle -\Delta_w u, u \rangle = \int (\widehat{-\Delta_w u})(\xi)\overline{\hat{u}(\xi)}d\xi = \int \hat{u}(\xi) \sum_{j=0}^{\infty} 2^{2j}|\hat{\psi}(2^{-j}\xi)|^2 \overline{\hat{u}(\xi)}d\xi =$$

$$= \int |\hat{u}(\xi)|^2 \sum_{j=0}^{\infty} 2^{2j}|\hat{\psi}(2^{-j}\xi)|^2 = |u|_B^2.$$

The same argument applies to the Fourier series of functions defined periodically on $[0,1]^N$. In particular, the fact that $\langle -\Delta_w u, u \rangle = |u|_B^2$ implies that the "wavelet Allen–Cahn" equation [10] with the translation invariant wavelet Laplacian

$$u_t = \epsilon \Delta_w u - \frac{1}{\epsilon}W'(u), \quad W(u) = (u^2 - 1)^2 \qquad (WAC)$$

describes the gradient descent in the problem of minimizing the WGL energy with the translation invariant Besov seminorm.

The steady states of (WAC) satisfy the respective Euler–Lagrange equation:

$$\epsilon \Delta_w u - \frac{1}{\epsilon}W'(u) = 0. \qquad (EL)$$

Convention Referring to the weak solutions of the gradient descent (wavelet Allen–Cahn) and Euler–Lagrange equations, we imply solutions to the following weak formulations. Function $u(x,t) \in H^1(\Omega \times [0,T])$ is a weak solution to the gradient descent equation for the WGL energy if

$$\int_0^T \langle u_t, v \rangle dt = \epsilon \int_0^T \mathcal{B}(u(\cdot,t), v(\cdot,t))dt - \frac{1}{\epsilon}\int_0^T \langle W'(u), v \rangle dt \text{ for any } v \in H^1(\Omega \times [0,T]).$$
$$(WAC_{weak})$$

Function $u(x) \in H^1(\Omega)$ is a weak solution to the Euler–Lagrange equation for the WGL energy if

$$\epsilon \mathcal{B}(u,v) - \frac{1}{\epsilon}\langle W'(u), v \rangle = 0 \text{ for any } v \in H^1(\Omega). \qquad (EL_{weak})$$

Here $\mathcal{B}(\cdot, \cdot)$ is the bilinear form defined on functions from H^1 as follows:

$$\mathcal{B}(f,g) = \langle f, \Delta_w g \rangle \text{ for any } f \in H^1, g \in H^2$$

and can be defined as

$$\mathcal{B}(f,g) = \lim_{n \to \infty} \langle f, \Delta_w g_n \rangle,$$

where $\{g_n\} \subset H^2$ is an arbitrary sequence approximating g in H^1. One can show that $B(f,g) \lesssim |f|_{H^1} \cdot |g|_{H^1}$ and the above limit indeed defines a bilinear form on H^1

(a similar argument is described in [11]). In the same sense as the wavelet Laplacian resembles the classical Laplace operator, the bilinear form is similar to the product $\int \nabla f \nabla g dx$. We use the notation $\mathcal{B}(\cdot, \cdot)$ due to the absence of a short convenient definition of a "wavelet gradient" (which would be possible if we used a completely discrete non-redundant and not translation invariant wavelet transform for the definition of the wavelet-based pseudo-differential operators and the Besov seminorm).

Our next step is to show that the above gradient descent problem with an appropriate initial condition is well-posed in the following sense: given any initial condition $u(x,0) \in H^1([0,1]^N)$ the gradient descent equation has a unique weak solution $u(x,t) \in H^1([0,1]^N \times [0, +\infty))$ that converges in $H^1([0,1]^N)$ to a steady state $u^*(x)$; this steady state is an extremum of the WGL energy and $WGL(u^*) < WGL(u)$ (in fact, $WGL(u(\cdot, t))$ is monotonely decreasing as $t \to \infty$).

Even though the wavelet Laplacian inherits many properties of the classical differential operator, as do the WGL and GL energies, the "wavelet Allen-Cahn" equation, unlike its classical analogue, does not satisfy a maximum principle (as it contains a non-local wavelet operator). Thus, in the following section we will often use energy methods that do not rely on maximum principle.

2.2. Gradient descent minimization of the WGL energy in the case of a regular wavelet. The following discussion of the gradient descent minimization of the WGL energy involves tools typical in the context of partial and ordinary differential equations. Those include, in particular, the Picard theorem for ODE in Banach spaces [14] and energy methods [16].

Assume that the wavelet that is chosen to construct the Besov seminorm in the definition of the WGL energy is regular, so that the associated space B is equivalent to H^1 ([17], [18]).

Let us investigate properties of the solution of the gradient descent equation arising in the problem of the WGL energy minimization.

$$\begin{cases} u_t = \epsilon \Delta_w u - \frac{1}{\epsilon} u^3 + \frac{1}{\epsilon} u, \\ u(x,0) = u_0(x) \in H^1([0,1]^N). \end{cases} \quad (GD)$$

Here Δ_w is the wavelet Laplace operator defined by (WvL). The following lemma gives some a priori regularity estimates for the solution of (GD). The norms and seminorms mentioned in the lemma are associated with the spaces of functions defined on $[0,1]^N$, so we omit the domain indication to simplify the notations(i.e. L^2 means $L^2([0,1]^N)$, etc), unless the domain is different from $[0,1]^N$ (in particular, $\mathcal{U} = [0,1]^N \times [0,T]$); the same applies to the domains of integration.

LEMMA 2.1. *Let $u(x,t)$ be a solution of the system (GD) on $\mathcal{U} = \{(x,t) : x \in [0,1]^N, t \in [0,T]\}$ with the initial condition $u(x,0) = u_0(x) \in H^1([0,1]^N)$ and let $E(u) = WGL(u)$. Then:*
(i) $\|u(\cdot, t)\|_{L^2}$, $\|u(\cdot, t)\|_{L^4}$, $|u(\cdot, t)|_{H^1}$ are uniformly bounded with respect to t;
(ii) $u(x,t) \in H^1(\mathcal{U})$ with $|u|^2_{H^1(\mathcal{U})} \leq E(u_0)(1 + \frac{c}{\epsilon})$, $c > 0$, the estimate is uniform with respect to T.

Proof. (i) The gradient descent system implies

$$\frac{d}{dt}\|u\|^2_{L^2} = -\epsilon |u|^2_B - \frac{1}{\epsilon}\int u^4 + \frac{1}{\epsilon}\int u^2 dx \leq -\epsilon |u|^2_B - \frac{1}{\epsilon}\left[\left(\int u^2\right)^2 - \int u^2\right].$$

Thus, $\frac{d}{dt}\|u\|^2_{L^2}$ becomes negative whenever $\|u\|_{L^2} > 1$, hence,

$$\|u\|_{L^2} \leq \max\{\|u_0\|_{L^2}, 1\}.$$

7

The WGL energy evaluated at the function $u(x, t)$ is non-increasing w.r.t. the time variable t:

$$\frac{d}{dt}E(u(x,t)) = \int u_t(-\epsilon\Delta_w u + \epsilon^{-1}(u^3 - u))dx = -\int(-\epsilon\Delta_w u + \epsilon^{-1}(u^3 - u))^2 dx \leq 0.$$
(2.1)

Hence, at all times $t \in [0, T]$ when the solution of the gradient descent equation exists, values $|u|_B^2$ and $\int u^4$ stay uniformly bounded:

$$|u|_B^2 \leq \frac{2}{\epsilon}E_0, \quad \|u\|_{L^4}^4 \leq 4\epsilon E_0 + 2\max\|u\|_{L^2}^2 \leq 4\epsilon E_0 + 2\max\{\|u_0\|_{L^2}, 1\}.$$

Therefore, (i) is true.

(ii) We notice that $\|u\|_{H^1(\mathcal{U})}^2 = \int_0^T \|u(\cdot, t)\|_{L^2}^2 dt + \int_0^T |u(\cdot, t)|_{H^1}^2 dt + \int_0^T \int u_t^2 dx$. The minimized energy stays non-negative, hence

$$\int_0^T \int u_t^2 dt \leq E(u_0)$$

for an arbitrarily large $T > 0$. This, together with the uniform bound on the norm of the solution u in $H^1([0, 1]^N)$, implies that $u(x, t) \in H^1(\mathcal{U})$ with the respective seminorm estimate. $\square$

Now, let us prove that the gradient descent problem associated with the minimization of the WGL energy is well-posed.

REMARK 1. *Since both the usual and translation invariant Besov $B_{2,2}^s$ seminorms are equivalent to the H^s seminorm (due to our assumption about the wavelet regularity), then for any function $u \in V_J$ one has $|u|_{B_{2,2}^s} \leq A_s 2^{Js}\|u\|_{L^2}$. In the lemma 2.2 we will use this fact with $s = 2$.*

LEMMA 2.2. *Let $B_R(0)$ denote an open ball of radius $R \geq 1$ centered at 0 in $L^2([0, 1]^N)$. The gradient descent equation projected on V_J with the initial condition $u(x, 0) = Pr_{V_J}u_0(x), u_0 \in B_R(0)$ has a unique solution in $V_J \cap B_R(0)$ which exists globally in time. Moreover, $u(x, t)$ converges to a solution of the corresponding Euler–Lagrange equation projected on V_J.*

Proof. Consider the projection of (GD) on a subspace V_J (the orthogonal projection on the subspace of the multiresolution analysis, via the discrete wavelet decomposition):

$$\partial_t Pr_{V_J} u = \epsilon Pr_{V_J}\Delta_w u - \frac{1}{\epsilon}Pr_{V_J}u^3 + \frac{1}{\epsilon}Pr_{V_J}u.$$

Since we are looking for a solution $u(x, t)$ that belongs to V_J w.r.t. x at any t, the Cauchy problem becomes

$$\begin{cases} u_t = \epsilon P_{V_J}\Delta_w u - \frac{1}{\epsilon}P_{V_J}u^3 + \frac{1}{\epsilon}u, \\ u(x, 0) = Pr_{V_J}u_0(x). \end{cases} \quad (GD_{V_J})$$

This gradient descent setting is associated with the minimization of WGL on V_J.

Let F denote the r.h.s. of the above equation:

$$F(u) = \epsilon Pr_{V_J}\Delta_w u - \frac{1}{\epsilon}Pr_{V_J}u^3 + \frac{1}{\epsilon}u.$$

The above ODE is autonomous, so it suffices to prove that F is locally Lipschitz. Indeed, let us consider the difference of the values of F at two elements $u, v \in V_J$ that

also belong to a fixed ball $B_R(u_0) \subseteq B_1(0)$

$$\delta F := F(Pr_{V_J}u) - F(Pr_{V_J}u) = \epsilon Pr_{V_J}\Delta_w(u - v) - \frac{1}{\epsilon}Pr_{V_J}(u^3 - v^3) + \frac{1}{\epsilon}P_{V_J}(u - v).$$

We will estimate the terms in the above sum separately. Using the fact that the Besov seminorms defined via the discrete and continuous wavelet transforms are equivalent and some more subtle estimates (given in detail in [12]) one can see that for any $f \in V_J$

$$c_1\|\Delta_w^d f\|_{L^2} \leq \|Pr_{V_J}\Delta_w f\|_{L^2} \leq c_2\|\Delta_w^d f\|_{L^2}, \text{ where } c_1, c_2 \text{ are independent of } f.$$

Finally, we get the following estimates for each term in the sum:

$$\|Pr_{V_J}\Delta_w(u - v)\|_{L^2} \leq a2^{2j}\|u - v\|_{L^2}, \; a = A_2 \text{ - the seminorms equivalency constant,}$$

$$\|u\|_{L^\infty} \leq C_{J,\alpha}2^{2(1+\alpha)J}\|u\|_{L^2} \text{ for any } \alpha > 0 \Rightarrow$$

$$\max\{\|u\|_{L^\infty}, \|v\|_{L^\infty}\} \leq C_{J,\alpha}2^{2(1+\alpha)J}\|u_0\|_{L^2}^2,$$

$$\|u^3 - v^3\|_{L^2}^2 \leq (3\max\{\|u\|_{L^\infty}, \|v\|_{L^\infty}\}^2)^2)^2\|u - v\|_{L^2}^2.$$

Therefore,

$$|\delta F|_{L^2} \leq (\frac{1}{\epsilon} + a\epsilon2^{4J} + 9C_{J,\alpha}^2 2^{4(1+\alpha)J}R^4)\|u - v)\|_{L^2},$$

here the Lipschitz constant $K = \frac{1}{\epsilon} + a\epsilon2^{4J} + 9C_{J,\alpha}^2 2^{4(1+\alpha)J}R^4$ depends on the radius R, parameter ϵ and the scale J of the multiresolution subspace V_J, all of which we fixed in advance, and is independent of the choice of the initial condition u_0.

Thus, by the generalization of the Picard theorem for Cauchy problems on a Banach space, there exists a time interval $[0, t_0]$, where $0 < t_0 \leq R/K$, on which (GD_{V_J}) has a unique solution $u(x, t)$.

Our next step is to show that the solution $u(x, t)$ stays within $B_R(0)$ at any $t \in [0, t_0]$, which will allow us to conclude that the solution can be extended infinitely in time.

After multiplying both sides of the projected (GD) equation

$$u_t = \epsilon Pr_{V_J}\Delta_w u - \frac{1}{\epsilon}Pr_{V_J}u^3 + \frac{1}{\epsilon}u$$

by u and integrating over the spatial domain, we get

$$\int uu_t = -\epsilon|u|_B^2 - \frac{1}{\epsilon}\int uPr_{V_J}u^3 + \frac{1}{\epsilon}\int u^2,$$

$$1/2\frac{d}{dt}|u|_{L^2}^2 = -\epsilon|u|_B^2 - \frac{1}{\epsilon}\int u^4 + \frac{1}{\epsilon}\int u^2,$$

$$1/2\frac{d}{dt}|u|_{L^2}^2 \leq -\epsilon|u|_B^2 - \frac{1}{\epsilon}(\|u\|_{L^2}^4 - \|u\|_{L^2}^2).$$

The right hand side of the above inequality becomes negative if either $\|u\|_{L^2([0,1]^N)}$ exceeds 1 or $\|u\|_{L^2([0,1]^N)} = 1$, $u \neq const$. Therefore, $u(x,t)$ stays within $B_R(0)$ when $t \in [0, t_0]$. Applying the same reasoning to the gradient descent system with an updated initial condition $u(x, t_0)$ we obtain the existence of solution for $t \in [0, 2t_0]$. Repeating the same procedure proves that the solution can be extended globally in time.

Now, since $\int u_t^2 dx dt$ is bounded (see proof of Lemma 2.1, part(ii) — namely, $\int_0^t u_t^2(x,s) dx ds \leq (1 + \frac{c}{\epsilon})E(Pr_{V_J}u_0)$ for any t),

$$\int u_t^2 dx \to 0 \text{ as } t \to \infty \text{ i.e. } u_t(x,t) \to 0 \text{ as } t \to \infty \text{ in } L^2([0,1]^N).$$

Hence, the limit of the solution $u_J^*(x) = \lim_{t \to \infty} u(x,t)$ satisfies the Euler–Lagrange equation for minimizers of WGL on V_J. Passing to the limit in both sides of the gradient descent equation is justified, since $u(\cdot, t) \in V_J$ and the wavelet ψ is sufficiently regular by assumption. $\square$.

THEOREM 2.3. *When $1 \leq N \leq 4$ the gradient descent equation with an initial condition $u(x,t) = u_0(x) \in H^1([0,1]^N)$ has a unique weak solution, which exists globally in time and satisfies $\|u\|_{L^2} \leq \max\{\|u_0\|_{L^2}, 1\}$.*

Furthermore, $u(x,t)$ converges to a steady state $u^*(x)$ in $H^1([0,1]^N)$ as $t \to \infty$, and $u^*(x)$ is a weak solution of the respective Euler–Lagrange equation

$$\epsilon \Delta_w u^* - \frac{1}{\epsilon}(u^*)^3 + \frac{1}{\epsilon}u^* = 0,$$

and thus is an extremum of WGL with a lower energy value than the one at the initial condition u_0.

Proof. Consider the sequence $\{u_n(x,t)\}$ of solutions to the following systems:

$$\begin{cases} u_t = \epsilon Pr_{V_n} \Delta_w u - \frac{1}{\epsilon} Pr_{V_n} u^3 + \frac{1}{\epsilon}u, \\ u(x,0) = Pr_{V_n} u_0(x). \end{cases} \qquad (GD_{V_n}^*)$$

Fix any finite $T > 0$. The WGL energy of each u_n decreases with time, since the ODE in $(GD_{V_n}^*)$ is a gradient descent equation for minimization of WGL on V_J. Therefore, the norms of u_n in $H^1([0,1]^N, [0,T])$ for any $T > 0$ are uniformly bounded, and hence, there exists a subsequence $\{u_{n_k}\}_{k=1}^\infty$ that converges weakly to a limit $u \in H^1([0,1]^N, [0,T])$.

We would like to conclude that the weak limit u solves the original gradient descent problem in the weak formulation, i.e. (WAC_{weak}) introduced in Section 2.1. Convergence of the time derivative and the wavelet Laplace (i.e. the bilinear form) component follows explicitly from the weak convergence of u_n to u in H^1. The proof of convergence for the non–linear part of the equation requires additional reasoning. In the case $N = 1$ there exists a uniform L^∞ bound for all u_n that guarantees the convergence of the respective cubes u_n^3 weakly in H^1:

$$\left| \int (u_n^3 - u^3) v \, dx dt \right| = \left| \int (u_n - u)(u_n^2 + u_n u + u^2) v \, dx dt \right| \leq$$

$$\leq \|u - u_n\|_{L^1} \cdot 3 \sup_n \|u_n\|_{L^\infty}^2 \|v\|_{L_\infty} \overset{n \to \infty}{\to} 0.$$

10

In cases $N > 1$ the sequence u_n converges to u in $L^p([0,1]^N, [0,T])$, $1 \le p < \frac{2(N+1)}{N-1}$. Indeed, noticing that

$$\int |u^3 - u_n^3| \le \left[\int |u - u_n|^p\right]^{1/p}\left[\int |u_n^2 + u_n u + u^2|^{p/(p-1)}\right]^{(p-1)/p},$$

we see that for any $N : 1 < N \le 4$ there exists such an exponent p that both p and $2p/(p-1)$ do not exceed $\frac{2(N+1)}{N-1}$. Therefore, for all such N it is true that u_n^3 converges to u^3 in L^1.

So, we conclude that u solves (GD) in the weak sense.

Let us prove that solution $u(x,t)$ of the (GD) problem converges to an extremum of WGL, i.e. to a steady state that weakly solves the Euler–Lagrange equation, as $t \to \infty$. Since $u_t = \epsilon\Delta_w u - \frac{1}{\epsilon}u^3 + \frac{1}{\epsilon}u$, the WGL energy evaluated at $u(\cdot, t)$ is non-increasing as $t \to \infty$, as was shown in the proof of Lemma 2.1 (equality (2.1)). Since the energy is also non-negative, the values $WGL(u(\cdot, t))$ converge to some value E_0, while $\frac{d}{dt}WGL(u(\cdot, t)) \to 0$. This implies convergence of $\epsilon\Delta_w u - \frac{1}{\epsilon}(u^3 - u)$ (and hence, of u_t) to 0 as $t \to \infty$ in the L^2 sense. At the same time, given any sequence of time points $\{t_n\}_{n=0}^\infty$ such that $t_n \to +\infty$, the sequence of functions $u(\cdot, t_n)$ has to have a subsequence converging to some function $u^* \in H^1([0,1]^N)$ (by the same compactness argument as in Theorem 3.5). Since solution u of the (GD) system is continuous with respect to t(as a function $[0,T] \to H^1([0,1]^N)$), after possibly being redefined on a set of measure 0), we conclude that $u(\cdot, t) \to u^*$ in $H^1([0,1]^N)$ as $t \to \infty$. An argument analogous to the one showing that u solves the gradient descent equation in the weak sense, one can prove that u^* satisfies the Euler–Lagrange equation for minimizers of WGL

$$\epsilon\Delta_w u^* - \frac{1}{\epsilon}((u^*)^3 - u^*) = 0$$

weakly. Therefore, u^* is an extremum point of WGL with a lower energy value than the energy at the initial function u_0. □

Let us remark that the proof of the existence of the (GD) solutions can also be deduced from Brezis' theory for monotone operators [5]. We thank one of the paper's reviewers for pointing this out.

2.3. General properties of WGL minimizers.

2.3.1. Geometric properties: diffuse interface analogy. The WGL energy was designed to match the structure of the diffuse interface models from material science and fluid dynamics. It also inherited an important property of the classical Ginzburg–Landau energy: when B seminorm is generated using a sufficiently regular wavelet, the WGL energy converges to its anisotropic sharp-interface limit as the scaling parameter ϵ tends to 0. The complete proof of this fact and description of the way the limiting functional depends on the chosen wavelet are given in [12]. Let us only remark that one of the main steps in the above analysis was to notice that for the binary functions ($u = \chi_E$, where E has a finite perimeter), the limit $\lim_{n\to\infty} \frac{|u_n|_B}{|u_n|_{H^1}}$ is independent of the choice of the approximating sequence $\{u_n\} \subset H^1$, $u_n \to u$ in L^2.

Later we will show how the value of ϵ influences the distribution of the Besov seminorm energy of a minimizer between wavelet scales. We will use the following notation for the scale associated with ϵ.

CONVENTION Let $J_\epsilon = min\{j : 2^{-j} < \epsilon\}$. For any $\epsilon > 0$ we separate all wavelet modes in two categories: high frequency modes $HF = \{\psi_{j,k} : j > J_\epsilon\}$ and low

frequency modes $LF = \{\psi_{j,k} : j \leq J_\epsilon\}$. This separation is an analogue of the notion "scale of the problem" used in the research of diffuse interface models. Later on in this paper projections of a function on the respective subspaces are called the *high and low frequency components* of this function.

The WGL minimizer existence and non-uniqueness follows from the fact that, in the case of the finite periodic domain, in particular, two minimizers of WGL are equal to the constant functions 1 and -1. If we take B to be the space of functions with finite B - seminorm that are defined on the entire space, it makes sense to take the double–well potential in a slightly adjusted form: $W(u) = u^2(u-1)^2$ (hence, $u = 0$ is a minimizer). In the most general case one can always use the compactness argument of Theorem 3.5. The non-uniqueness also arises from the energy being even w.r.t. u.

The following lemma explains the nature of the separation of the wavelet scales into the low and high frequencies and the role of the diffuse interface parameter ϵ in this definition.

LEMMA 2.4. *If u solves (EL), its low frequency components are dominating:*

$$\sum_{j=0}^{J_\epsilon}(1 - 2^{2j}\epsilon^2) \int c_{j,k}^2 dk > \sum_{j=J_\epsilon+1}^{\infty}(-1 + 2^{2j}\epsilon^2) \int c_{j,k}^2 dk$$

and

$$\sum_{j=0}^{J_\epsilon} \int c_{j,k}^2 dk > 2 \sum_{J=J_\epsilon+1}^{\infty}\sum_{j=J}^{\infty} \int c_{j,k}^2 dk$$

where $c_{j,k} = \langle u, \psi_{j,k}\rangle$, $j = 0, 1, ..., k \in \mathbb{R}^N$.

Proof. (EL) implies

$$\langle u^3, \psi_{j,k}\rangle = (1 - \epsilon^2 2^{2j})c_{j,k},$$

hence, for all $j > -\log_2 \epsilon$ the solution u and its cube u^3 have either coefficients of opposite sign, or both zero coefficients. Then

$$\int u^3 = \int u, \quad \sum_{j=0}^{J_\epsilon}(1 - 2^{2j}\epsilon^2)c_{j,k}^2 > \sum_{j=J_\epsilon+1}^{\infty}(-1 + 2^{2j}\epsilon^2)c_{j,k}^2,$$

$$\sum_{j=0}^{J_\epsilon} \int c_{j,k}^2 dk > 3 \sum_{J=J_\epsilon+1}^{\infty}\sum_{j=J}^{\infty} \int c_{j,k}^2 dk.$$

The last equality implies

$$\sum_{j=0}^{J_\epsilon} \int c_{j,k}^2 dk > (2^{2(J-J_\epsilon)} - 1)\sum_{j=J}^{\infty} \int c_{j,k}^2 dk,$$

and, moreover,

$$\sum_{j=0}^{J_\epsilon} \int c_{j,k}^2 dk > 2 \sum_{J=J_\epsilon+1}^{\infty}\sum_{j=J}^{\infty} \int c_{j,k}^2 dk.$$

12

⬚

In the case of a regular wavelet ψ generating the classical Besov seminorm the above lemma implies that ϵ controls the steepness of transitions that the minimizer is allowed to have.The smaller is ϵ, the more wavelet scales are included in the dominating low frequency (LF) range, thus allowing the minimizer to change faster.

Fig.2.2 shows two extrema obtained numerically by solving the gradient descent system (with the same initial condition) to steady state. Solutions (a) and (b) correspond to the problems with two different values of parameter ϵ: $1/2$ and $1/16$ respectively.

(a)　　　　(b)
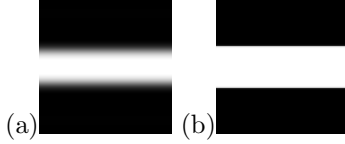
FIG. 2.2. *Steady state solutions of the gradient descent equation with (a) $\epsilon = 1/2$, (b) $\epsilon = 1/16$.*

Remark. Before we continue discussing the minimizers, let us remark that in this paper there are two different situations when we talk about the value of epsilon in relation to the minimizer properties. The first is the description of the dependence between the value of ϵ and the "diffuse interface width", which, in our case is expressed in terms of how many coarsest wavelet levels contain the dominating part of the Besov-seminorm energy (lemma 2.4 above). Another situation is the study of the modified WGL energy with additional forcing terms, where we show that by choosing the forcing term weights large enough (comparing to ϵ^{-1}) we can achieve the minimizer uniqueness. In the applications it makes sense to choose the value of ϵ depending on how sharp we require the edges of the output to be, and how large are the gaps we might need to connect those edges over (inpainting case). After that we can choose the values of the forcing parameters - depending on how important it is for us to preserve certain spatial and/or edge-related info.

2.3.2. More on regularity of WGL minimizers.

LEMMA 2.5. *Any solution of (EL) that belongs to L^∞ is also infinitely smooth.*

The proof of a more general fact can be found in [12].

REMARK (EL) has no non-trivial binary solutions. Indeed, if u is a steady state and $|u(x)| = 1$ a.e., then $\Delta_w u(x) = 0$, hence, u is trivial. Hence, (GD) has no non-trivial binary steady states. The results of the binary image recovery that we consider in the next section are grayscale rather than binary. However, those are sufficiently close to the binary class so that simple thresholding procedure returns a binary result close to the original solution.

This also implies that the WGL minimizers can have no jumps between two exact steady states of the double–well. Non-constant minimizers have to assume other, "grayscale" values between -1 and 1, which means that the model indeed has the "diffuse interface" nature.

LEMMA 2.6. *Every solution of (EL) on $\mathbb{R}^N$ belongs to the Sobolev $W^{2,1}$ space.*

Proof. If u be a solution of (EL), then $\Delta_w u = \frac{1}{\epsilon^2}(u^3 - u)$. Since $u \in L^4$ (WGL energy assumes a finite value at u), $\Delta_w u \in L^1(\mathbb{R}^N)$. So, by the Hausdorff-Young inequality,

$$\hat{u}(\xi) \sum_{j=0}^{\infty} 2^{2j}|\hat{\psi}(2^{-j}\xi)|^2 \leq \int |(u^2(x) - 1)u(x)|dx \leq \|u\|_{L^2}\|u^2 - 1\|_{L^2}.$$

13

Now, since

$$\sum_{j=0}^{\infty} 2^{2j} |\hat{\psi}(2^{-j}\xi)|^2 = O(|\xi|^2), \ |\xi| \to +\infty,$$

any of the second order derivatives of u is measurable, with its Fourier transform in L^∞, hence, it belongs to L^1. □

LEMMA 2.7. *Every minimizer of WGL in one or two dimensions*

$$WGL(u) = \frac{\epsilon}{2}|u|_B^2 + \frac{1}{4\epsilon} \int W(u(x))dx$$

belongs to L^∞.

Proof. 1. For $N = 1$ the statement follows from the Sobolev embedding $H^1 \subset\subset L^\infty$.
2. Let $N = 2$. Every minimizer of WGL satisfies

$$\epsilon \Delta_w u = \frac{1}{\epsilon}(u^3 - u) = 0.$$

After differentiating both sides of the above equality one can express a derivative of $\Delta_w u$ as :

$$\epsilon^2 \partial_x \Delta_w u = \partial_x u(3u^2 - 1),$$

and see that its L^1 norm is finite

$$\epsilon^2 \int |\partial_x \Delta_w u| \le |u|_{H^1} \Big[\int (3u^2 - 1)^2\Big]^{1/2} < \infty.$$

Therefore, $\Delta_w u \in W^{1,1}$ and $\xi \widehat{\Delta_w u}(\xi) \in L^\infty$. Since we know $\widehat{\Delta_w u}(\xi) = \hat{u}(\xi)O(|\xi|^2)$, $\xi \to \infty$ we conclude $\hat{u}(\xi) \le \frac{C_0}{|\xi|^3}$, hence, $\hat{u}(\xi) \in L^1$ (or l^1 in case of the periodized domain/Fourier series) and $u \in L^\infty$. □

COROLLARY 2.8. *Every minimizer of WGL in one or two dimensions is infinitely smooth.*

This statement follows directly from the lemmas 2.5 and 2.7.

3. Adding the fidelity terms in the spatial and wavelet domains. The examples we describe below generalize our previous work on WGL inpainting. The modified WGL energy we describe below is designed for the general problem of image recovery. We include an additional fidelity term involving wavelet coefficient data in the model together with the spatial fidelity. The resulting energy can be used for a variety of image recovery applications and serves as an especially convenient tool for applications without clearly defined problem type, i.e. inpainting with undetermined missing area, superresolution of noisy images, combined inpainting and superresolution and so on.

3.1. Modified WGL energy with fidelity terms (WGLF). In the most general form, the modified WGL energy that we are going to discuss can be defined as a sum of the WGL functional itself and two additional fidelity terms — the edge–preserving and the spatial consistency term:

$$E(u) = WGL(u) + \frac{\mu}{2}\|(u - f)\chi_\Omega\|_{L^2}^2 + \frac{\lambda}{2}|P_\Lambda(u - f)|_B^2,$$

14

χ_Ω and χ_Λ are indicators of the information to be preserved in the spatial and wavelet domains respectively. We assume that the set of preserved modes is finite so $|P_\Lambda(u - f)|_B^2$ is a valid expression, despite the fact that the original function f might not belong to B (for instance, if ψ is regular and f is binary). This model can be used in a variety of applications, and is adaptable to each of those by a special choice of the forcing term(s) and the initial guess.

The spatial fidelity term is designed to preserve the truly known values of the function in case those are given. The spatial mask is the indicator function of a subset of the function domain where the function values are known. This term can be assigned a weight μ depends on the importance of the spatial constraint and the quality of given data (μ can be increased in cases when the given image has high contrast and low noise level). The edge-forcing mask χ_Λ is defined on the set of all wavelet indices, it is simply the indicator function of Λ — an a priori chosen set of wavelet modes.

In general, the choice of forcing terms and their weights varies depending on each particular application and properties of the initially known data. More detailed discussion of the properties of both fidelity terms shows that in many cases those terms, being present exclusively, produce almost identical steady states in the case of the binary inpainting problem with a predefined unknown area. Nevertheless, the edge forcing term allows to use the WGLF minimization method (with the appropriate choices of Ω and Λ) for a much wider set of image reconstruction problems.

The existence of minimizers of WGLF can be proven in the same way as for the unmodified WGL energy: indeed, adding two more non-negative terms to the original WGL functional brings no change to the compactness argument used to prove Theorem 3.5.

3.2. Variational properties of the modified WGLF energy with the L^2 spatial fidelity term. We will prove several lemmas characterizing minimizers of the modified WGL energy that includes only the spatial fidelity:

$$E(u) = WGL(u) + \frac{\mu}{2}\|(u - f)\chi_\Omega\|_{L^2}^2,$$

We will abbreviate this form of modified WGL energy as WGLFs.

LEMMA 3.1. *Let u and v be two solutions of the Euler–Lagrange equation for the modified WGL energy with the spatial fidelity ($\lambda = 0$) and let $\mu > \frac{1}{\epsilon}$. Any of the following conditions is sufficient for the identical equality $u \equiv v$:*
1. minimizers coincide on the unknown domain: $\int_{\Omega^c}(u - v)^2 = 0$;
2. the L^2-norm of the difference on the unknown domain is significantly smaller than the same on the known domain: $\int_{\Omega^c}(u - v)^2 < (\mu\epsilon - 1)\int_\Omega(u - v)^2$;
3. low frequency components of u and v are the same: $|P_{LF}(u - v)|^2 = 0$.

Proof. Both u and v satisfy (EL):

$$\epsilon\Delta_w u - \frac{1}{\epsilon}(u^3 - u) - \mu\chi_\Omega(u - f) = 0 \text{ and } \epsilon\Delta_w v - \frac{1}{\epsilon}(v^3 - v) - \mu\chi_\Omega(v - f) = 0,$$

Hence, after multiplying the difference of the above equations by $(u - v)$ and integration one gets

$$\epsilon|u - v|_B^2 + \frac{1}{\epsilon}\int (u - v)^2(u^3 + uv + v^3) - \frac{1}{\epsilon}\|u - v\|_{L^2}^2 + \mu\int_\Omega(u - v)^2 = 0.$$

Now, both condition 1 and 2 imply

$$\epsilon|u-v|_B^2 + \frac{1}{\epsilon}\int (u-v)^2(u^3 + uv + v^3) = 0,$$

which is only possible if $u \equiv v$. Condition 3 implies $\epsilon|u-v|_B^2 \geq \frac{1}{\epsilon}\|u-v\|_{L^2}^2$, which, by similar reasoning, occurs only if $u - v \equiv 0$. □

The following lemma describes a set of functions that contains exactly one minimizer of WGLF whenever $\mu\epsilon > 1$. This set includes all functions with sufficient portion of the "low frequency" components localized inside of the known domain Ω.

LEMMA 3.2. *Let $\mu\epsilon > 1$, $\lambda = 0$ then WGLF is a convex functional on the following subspace of functions:*

$$X = \{u \in B : (\mu\epsilon - 1)\int_\Omega u_{LF}^2 \geq \int_{\Omega^c} u_{LF}^2\}.$$

Consequently, if there exists a solution u to the Euler-Lagrange equation such that $u \in X$, then this solution is unique and it is the global minimizer of the WGLF energy.

Proof. We need to show that the second variation of WGLF is positive definite on X:

$$\epsilon|y|_B^2 + \frac{1}{\epsilon}\int ([3u^2 - 1]y^2)dx + \mu\int_\Omega y^2 > 0 \quad \forall y \in X \setminus \{\Theta\}.$$

To prove this condition it suffices to show that

$$\epsilon|y|_B^2 - \frac{1}{\epsilon}\int y^2 dx + \mu\int_\Omega y^2 > 0 \quad \forall y \in X \setminus \{\Theta\}.$$

Let us rewrite the l.h.s. of the above inequality by separating the LF and HF modes and grouping some terms:

$$\epsilon|P_{LF}y|_B^2 + (\epsilon|P_{HF}y|_B^2 - \frac{1}{\epsilon}\int P_{HF}y^2 dx) + (-\frac{1}{\epsilon}\int P_{LF}y^2 dx + \mu\int_\Omega P_{LF}y^2) + \mu\int_\Omega P_{HF}y^2.$$

Now, the second terms is non-negative by definition of the HF modes and the third terms is non-negative due to the property defining the functions in X, thus the entire sum is positive. Indeed, if $P_{LF}y \neq 0$, the first term is positive, and if $P_{LF}y = 0$, then, as $y \neq 0$, we have $P_{HF}y \neq 0$, and the second term in the sum is strictly positive.

Therefore, any solutions of the EL equation in this case have to be local minimizers of the energy, thus there exists at most one global minimizer. □

Let us also remark that X is a non-zero subspace, for instance, if $f(x) = 1$ for any $x \in \Omega$ and $(\mu\epsilon - 1)|\Omega| > |\Omega^c|$, in which case the unique minimizer is identically equal to 1.

COROLLARY 3.3. *Let $\mu\epsilon > 2$, and $|\Omega| > |\Omega^c|$ (unknown domain is smaller) then WGLFs has a unique minimizer among all functions with L^∞ norm not exceeding 1.*

Since we are looking for nearly-binary functions that solve the inpainting problem, this criterium gives a reasonable sufficient condition for the uniqueness of the minimizer that fits the role of a binary image.

The following corollary states that the minimizer uniqueness for the WGL energy defined using the non-redundant wavelet transform can be guaranteed by imposing

conditions on the wavelet basis functions and the inpainting (missing) domain geometry.

COROLLARY 3.4. *Let the B-seminorm in the definition of WGLF be defined via the non-redundant wavelet transform based on the orthonormal wavelet basis $\{\psi_{j,k}\}$. If any $\psi_{j,\kappa}$ with $j < -\log_2 \epsilon$ (any LF wavelet kernel value) satisfies*

$$(\mu\epsilon - 1)\int_\Omega \psi_{j,k}^2 > \int_{\Omega^c} \psi_{j,k}^2,$$

then the modified WGL energy with the spatial fidelity term has a unique minimizer in B.

3.3. Variational properties of the modified WGL energy with both fidelity terms. The proof of the minimizer existence for the WGLF energy uses the standard "compactness argument".

THEOREM 3.5. *The WGLF energy functional*

$$E(v) = \frac{\epsilon}{2}|v|_B^2 + \frac{1}{4\epsilon}\int W(v)dx + \frac{\mu}{2}\|(u-f)\chi_\Omega\|_{L^2}^2 + \frac{\lambda}{2}|P_\Lambda(u-f)|_B^2,$$

has at least one minimizer on B, elements of which can be defined either on a torus or on the entire space. The minimizer may not be unique.

Proof. The energy functional E is non-negative, and hence has a finite infimum $M < \infty$. Let us prove the infimum is attained on some function from the admissible set B. Consider a minimizing sequence $\{u_n\} : E(u_n) \to M$ as $n \to \infty$. The sequence of energy values $E(u_n)$ converges, and hence is bounded, and so are the seminorms $|u_n|_B$.

The corresponding norms $\|u_n\|_B$ can be estimated from above using the bound on the integral of the double–well:

$$\int W(u_n)dx \le C \Rightarrow |\int u_n^2 - 1| \le C'', \quad \int u_n^2 \le C'' + 1,$$

where C', C'' are positive constants. Thus, the sequence $\|u_n\|_B$ is also bounded:

$$\|u_n\|_B^2 \le |u_n|_B^2 + \|u_n\|_{L^2}^2,$$

and since B is Banach, $\{u_n\}$ has a weakly converging subsequence: $\{u_{n_k}\} : u_{n_k} \rightharpoonup u \in B$. The sequence is also bounded in L^4, so, WLOG we can assume that u_{n_k} converges weakly in L^4 as well. Fix some $v = u_{n_{k*}}$. Then

$$\frac{\epsilon}{2}\langle u_{n_k}, v\rangle_B + \frac{1}{4\epsilon}\int (u_{n_k}^2 - 1)(v^2 - 1)dx + \frac{\mu}{2}\langle u_{n_k} - f, v - f\rangle_{L^2(\Omega)}^2 +$$

$$+\frac{\lambda}{2}\langle P_\Lambda(u_{n_k} - f), P_\Lambda(v - f)\rangle_B^2 \to \frac{\epsilon}{2}\langle u, v\rangle_B + \frac{1}{4\epsilon}\int (u^2 - 1)(v^2 - 1)dx +$$

$$+\frac{\mu}{2}\langle u - f, v - f\rangle_{L^2(\Omega)}^2 + \frac{\lambda}{2}\langle P_\Lambda(u - f), P_\Lambda(v - f)\rangle_B^2.$$

Now, if we let $k^* \to \infty$, we conclude that

$$\frac{\epsilon}{2}\langle u_{n_k}, u_{n_{k*}}\rangle_B + \frac{1}{4\epsilon}\int (u_{n_k}^2 - 1)(u_{n_{k*}}^2 - 1)dx + \frac{\mu}{2}\langle u_{n_k} - f, u_{n_{k*}} - f\rangle_{L^2(\Omega)}^2 +$$

17

$$+\frac{\lambda}{2}\langle P_\Lambda(u_{n_k}-f), P_\Lambda(u_{n_{k^*}}-f)\rangle_B^2 \to E(u) \text{ as } k, k^* \to \infty.$$

Taking the diagonal subsequence with $k = k^*$ we see that $E(u_{n_k}) \to E(u)$ as $k \to \infty$. Hence, the minimum of the WGLF energy is achieved at the element $u \in B$. $\square$

The Euler–Lagrange equation for the modified WGL energy with both fidelity terms takes the form

$$\epsilon\Delta_w u - \epsilon^{-1}(u^3 - u) - \lambda\Delta_w P_\Lambda(u-f) - \mu\chi_\Omega(u-f) = 0.$$

Let us describe the critical points of the modified WGL energy based on the sign of its second variation. Second variation of WGLF at a point $v \in B$ is

$$\delta_v^2 WGLF(u) = \epsilon|v|_B^2 + \frac{1}{\epsilon}\int(3u^2-1)v^2dx + \lambda|P_\Lambda v|_B^2 + \mu\int_{\Omega^c} v^2dx.$$

The above expression cannot be negatively definite w.r.t. v, since it is positive for v from any subspace W_j where $j > -\log_2\epsilon$ independently of u. Therefore, *modified WGL functional does not have any local maxima.*

A natural question to ask is: in which cases is it possible to recover a function f from the variational WGLF model exactly? The following lemma describes one of the possible situations allowing for such recovery.

LEMMA 3.6. *If f is a local minimizer of the WGL energy, then $u = f$ is a local minimizer of the modified WGLF energy.*

Proof. Since f is a local minimizer of the WGL energy and the global minimizer of the fidelity terms, it is a local minimizer of the WGLF. $\square$ If, additionally, it is known that the WGLF has a unique minimizer, then the function f can be recovered.

Let us estimate some parts of the sum that represents $\delta_v^2 WGFL(u)$.

$$\delta_v^2 WGL(u) = \left[\epsilon|v_{HF}|_B^2 - \frac{1}{\epsilon}\int_{\Omega^c} v_{HF}^2dx\right] + (\mu - \frac{1}{\epsilon})\int_{\Omega^c} v^2dx +$$

$$+ \left[\lambda|P_\Lambda v|_B^2 - \frac{1}{\epsilon}\int_{\Omega^c} v_{LF\cap\Lambda}^2dx\right] + \left[\epsilon|v_{LF}|_B^2 - \frac{1}{\epsilon}\int_{\Omega^c} v_{LF\backslash\Lambda}^2dx + \frac{1}{\epsilon}\int 3u^2v^2dx\right].$$

The first term is non-negative by the definition of the HF components, the second and third terms are non-negative provided $\mu > \frac{1}{\epsilon}$ and $\lambda > \frac{1}{\epsilon}$ respectively. So, this representation allows one to formulate a sufficient condition of u being a local minimizer of the WGLF energy.

LEMMA 3.7. *Let u be a solution of the EL equation for WGLF with $\mu > \frac{1}{\epsilon}$ and $\lambda > \frac{1}{\epsilon}$. For u to be a local minimizer it suffices to satisfy*

$$\epsilon^2|v_{LF}|_B^2 - \int_{\Omega^c} v_{LF\backslash\Lambda}^2dx + \int 3u^2v^2dx > 0 \quad \forall v \in C_0^\infty.$$

The lemma itself does not give a clear intuitive understanding of the nature of this condition. However, its corollaries have a straightforward connection to the imaging problems. We are going to assume that $\frac{1}{\epsilon} < \lambda, \mu$ everywhere in the further discussion.

COROLLARY 3.8. *If $\Lambda \supseteq LF$, any solution of the EL equation for the modified WGL energy is a minimizer. Therefore, the minimizer is unique.*

In image processing terms, this defines the relation between the diffuse interface parameter ϵ and criteria for the known information to be sufficient to recover the

unique output. Namely, knowing all of the LF components of the function one can recover it in a unique way using the WGLF model. This applies in particular to the superresolution problem, which is described in detail in Section 4.2. Fig.3.1 shows the underresolved picture of a black triangle, the set of edges that were included in the forcing term and the unique output of the minimization algorithm. In this case we took $\epsilon = 1/16$ which corresponds to LF wavelet scales $j = 1, 2, 3, 4$. Having the values of the wavelet transform fixed at those scales, the WGLF minimization algorithm recovers one the same output (Fig.3.1 (c)) for any initial condition (with all other parameters being identical).

(a) (b) (c)

FIG. 3.1. *(a) 4 times underresolved image; (b) edges included in the forcing term of WGLF (based on the 'DB4' wavelet); (c) minimization output.*

COROLLARY 3.9. *If u is a solution of the EL equation for WGLF and $|u(x)| \geq 3^{-1/2}$ for a.e. $x \in \Omega^c$, then u is a local minimizer of WGLF. Therefore, the minimizer is unique on the set of functions*

$$\{v \in B : |v(x)| \geq 3^{-1/2} \text{ for a.e. } x \in \Omega^c\}.$$

This statement merely says that the WGLF minimization problem has a unique solution on the set of functions that do not have significant variations of color (including those that involve changing its sign) on the unknown domain Ω^c.

3.4. Edge preserving wavelet fidelity term. The following section discusses a possible construction of the edge–preserving fidelity term in the WGLF energy. The effectiveness of such fidelity term choice in practice is illustrated with several computational examples.

3.4.1. Thresholding of the wavelet coefficients for binary functions. Thresholding of wavelet coefficients has been efficiently used in image processing in a variety of ways (in particular, for the translation-invariant thresholding see [9]). The thresholding method described in this section is designed specifically with the goal of choosing the subset of wavelet modes Λ that constitute the forcing term of WGLF, thus allowing to enforce only those features that are considered significant for the given image and ignore edges introduced by any undesired occlusions.

Since any binary image is effectively an indicator function of a bounded measurable set, we would like to define a unified procedure for the classification of significant vs. not significant wavelet coefficients of a binary function. Then, after choosing all significant wavelet modes for the given binary image one can, using the same evaluation method, find all coefficients significant for the occlusion indicator and exclude those from the edge–preserving fidelity term. Here we account for two main factors influencing the thresholding parameters: the properties of the wavelet decomposition of binary functions (characteristic functions of finite perimeter sets) and the fact that the edge–preserving fidelity term penalizes the deviation of wavelet coefficients measured in terms of the Besov seminorm.

Let us introduce some general notations for the thresholding procedure. Given a function $u \in L^2$ we define an associated boolean function $b = b(u)$ to indicate the

significance of each coefficient with index $(j, \vec{k})$ in the entire wavelet decomposition

$$u(x) = \sum_{j, \vec{k}} c_{j, \vec{k}} \psi_{j, k}(x),$$

in accordance with a certain criterium $\mathcal{S}$. Namely, $b : \mathcal{I} = \{(j, \vec{k})\}_{j \in \mathbb{Z}, \vec{k} \in \mathbb{Z}^n} \mapsto \{0, 1\}$,

$$b(j, \vec{k}) = \begin{cases} 1, & \mathcal{S} \text{ is true for } c_{j, \vec{k}}, \\ 0, & \text{otherwise.} \end{cases} \qquad (THR)$$

In the inpainting context

$$\Lambda = \operatorname{supp} b \subset \mathcal{I}.$$

We describe possible choices of $\mathcal{S}$ based on the scale-dependent thresholding of wavelet coefficients, which can either be uniform within each scale or relative, estimating each coefficient as compared to the rest of the wavelet decomposition. Such criteria occur to be applicable to most of the images containing visually pronounced contours. Since all applications we consider deal with images in 2D, the spatial rescaling constants are those for the wavelet decomposition in $\mathbb{R}^2$.

As the thresholding procedure is entirely application-determined, we will consider functions from the spaces $\mathcal{V}_J$ and assume that their wavelet decomposition is translation invariant with respect to translations by multiples of 2^{-J}, i.e. translations at the finest wavelet scale, or one-pixel translations. In this manner, any wavelet decomposition of a function $u \in \mathcal{V}_J$ involves 2^{2J} coefficients at each of the wavelet scales.

ABSOLUTE (BESOV) THRESHOLDING Given a fixed thresholding parameter $\mathcal{T} > 0$, we define the **absolute significance threshold at scale j** as

$$\tau_j = 2^j \mathcal{T}.$$

Then for any function $u \in L^2$ the logical statement $\mathcal{S}$ in (THR) can be written as

$$(j, k) \to \Lambda \Leftrightarrow |\langle u, \psi_{j,k} \rangle| > \tau_j,$$

which, is equivalent to

$$(j, k) \to \Lambda \Leftrightarrow |\langle u, \psi_{j,k} \rangle \psi_{j,k}|_B > \mathcal{T}$$

in terms of the B -seminorm estimates.

While the classical hard thresholding uses a fixed threshold level, the above absolute thresholding is scale-dependent: we fix the uniform parameter, but apply the hard thresholding to each scale with the rescaled parameter (multiplied by 2^j).

For a reasonably chosen $\mathcal{T}$ the absolute Besov thresholding can be viewed as a denoising procedure, because one disregards all details that are below the significance threshold uniformly defined for each scale. The drawback of using the absolute thresholding is the necessity to set the thresholding parameter $\mathcal{T}$ in advance, moreover, the choice of this value may require preliminary testing under user supervision.

The need for non-uniform, "relative" thresholding arises from the necessity to automatize the procedure completely, as well as from the idea of using different thresholds for edges in separate directions.

RELATIVE THRESHOLDING Let us define the thresholding rule for each scale of wavelet decomposition of functions $u \in \mathcal{V}_J$. In order to do that, consider the expected value(the mean) M_j and the standard deviation D_j for the set of all wavelet coefficients $c_{j,k} = \langle \psi_{j,k}, u \rangle$ at scale $j \in \mathbb{N}$:

$$M_j = 2^{-Nj} \sum_{k_i=0}^{2^J-1} c_{j,k},$$

$$D_j^2 = 2^{-Nj} \sum_{k_i=0}^{2^J-1} (c_{j,k} - M_j)^2 = 2^{-Nj} \sum_{k_i=0}^{2^J-1} c_{j,k}^2 - M_j^2.$$

DEFINITION We define the **relative significance threshold at scale j** as

$$\tau_j = C 2^j D_j, \; C = 2^{-J_{max}},$$

where J_{max} is the **thresholding scale**, which is defined as the maximum level of wavelet decomposition J (i.e. the image resolution) or as the scale that stores the most significant information (visually significant, or defined by a specific application or the given data quality). In practice, to make the thresholding more precise, the coefficient C may be adjusted by an additional multiple reflecting the quota of the edges in the image.

In this manner, we define the following unified criterium $\mathcal{S}$ for the relative wavelet thresholding.

DEFINITION A mode $\psi_{j,\vec{k}}$ is chosen to be **relatively significant** for a function u, i.e. $b_u(j,k) = 1$ if and only if it differs from the mean coefficient value at the scale j by more than the standard deviation times the dyadic scaling multiple:

$$|\langle u, \psi_{j,k} \rangle - M_j| \geq C 2^j D_j.$$

The relative thresholding leaves intact those coefficients that differ sufficiently from the mean of all coefficients at this wavelet scale. Let us explain the reasons for such choice of the thresholding condition.

Consider a characteristic function of some measurable set $u = \chi_E \in L^2(\mathbb{R}^2)$. Due to u being locally homogeneous of degree 0, we get:

$$\langle u, \psi_{j,k} \rangle \asymp 2^p \langle u, \psi_{j+p, 2^p k} \rangle.$$

Only wavelet modes supported in the neighborhood of the set boundary ∂E may produce non-zero projections, thus the total amount of non-zero coefficients at scale j grows as $O(2^j)$. The decrease in the coefficient values as the scale increases and the respective increase of the integration domain for the translation parameter k makes the mean value of the non-zero coefficients stay within the same order $O(1)$, while the mean value of all coefficients decreases as $O(2^{-j})$. The standard deviation, for the same reasons, changes as $O(2^{-j})$ as the scale j increases. Therefore, the coefficients of wavelet modes supported near ∂E should differ from the mean value by a value comparable to $2^j D_j$. The exact choice of the coefficient C in practice depends on the importance of reconstructing an exactly "binary" function, the set geometry (approximate length of the boundary ∂E), etc.

Fig.3.2 shows an example of a binary function (the indicator of the set corresponding to the white part of the image), and the function reconstructed from the wavelet

coefficients that were not eliminated by relative thresholding. The thresholding scale in this setting equals the finest scale of possible decomposition for an image 256×256.
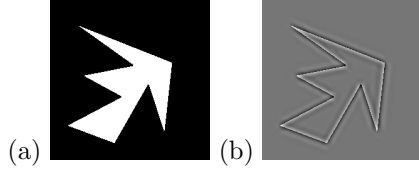

FIG. 3.2. *Example of the indicator function thresholding: (a) a binary image, (b) its edges reconstructed from "significant" wavelet coefficients of the stationary wavelet transform.*

The described technique itself can be useful in filtering out the major jumps in the signals that, in theory, are supposed to correspond to continuous functions ([13]).

3.4.2. Locally adaptive thresholding. Going back to the problem of image inpainting, let us recall that Ω denotes the part of the domain where function is defined (known), and Ω^c is the set where function values are missing (occluded, corrupted). We define the **locally adaptive thresholding** as a procedure of the simultaneous scale–dependent thresholding of the image u and the indicator function of its unknown area χ_{Ω^c} using the evaluation criterium $\mathcal{S}$ that is a conjunction of two statements: "the mode corresponds to a coefficient of u that is above significance threshold" and "the mode corresponds to a coefficient of χ_{Ω^c} that is below its significance threshold". In the case of absolute thresholding, it can be expressed as the following rule of choosing significant modes:

$$(j, \vec{k}) \in \Lambda \Leftrightarrow |\langle u, \psi_{j,k}\rangle| \geq 2^j T, \quad |\langle \chi_\Omega, \psi_{j,k}\rangle| \leq 2^j t,$$

where T and t are fixed significant thresholds for functions u and χ_Ω respectively. To use the relative thresholding in the same setting, one needs to replace t and T by the deviation–based parameters computed separately for each scale of the image. Fig.3.3 shows the result of locally adaptive relative thresholding. The binary (black and white) image has a missing area that is shown in gray. The r.h.s. image shows the function that was reconstructed from wavelet coefficients not eliminated by the relative locally adaptive thresholding, i.e. coefficients that are significant for the binary image and are not correlated with the occlusion. Getting somewhat ahead of the current discussion, Fig.3.3 (c) shows the result of binary inpainting preserving the edges shown in Fig.3.3 (b).
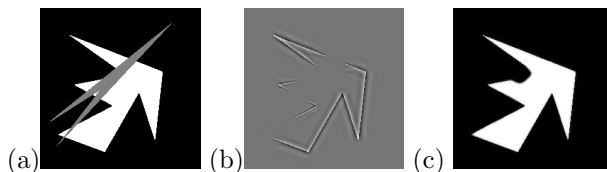

FIG. 3.3. *Example of the locally adaptive wavelet coefficient thresholding: (a) given binary image with the missing part indicated by gray, (b) the result of relative locally adaptive thresholding: the edges of the occlusion are ignored, while the edges of the original image are preserved, (c) the result of inpainting using WAC with the edge fidelity.*

22

The next section deals with the energy functional designed for the variational image inpainting. Constructing this functional involves the choice of the edges that one would like to be preserved in the process of inpainting, this is where the above thresholding methods are applied.

3.5. Modified WGL energy for image inpainting. This subsection describes the application of the introduced modified WGL energy model to image inpainting. The problem of image inpainting using the WGL energy with the spatial fidelity term has been addressed in [10]. Adding the edge–preserving fidelity allows to deal with certain marginal situations within the same class of inpainting problems, such as inpainting with unclearly defined missing area or direction-dominated inpainting.

3.5.1. Binary inpainting model with edge–preserving fidelity term. As described before, the inpainting problem set-up necessarily includes specification of the set Λ. We define this set so that it includes coefficients of the image at each scale that are relatively significant for the entire image, and are not relatively significant for the occlusion indicator (i.e. perform locally adaptive thresholding). The knowledge of relatively significant coefficients of the mask χ_{Ω^c} allows to perform **locally adaptive thresholding**: to choose the set of wavelet coefficients of function u that are minimally correlated to the mask.

The most natural way to approach the inpainting problem is to use a simple technique to preprocess the image and apply a more computationally expensive one to refine it. Polynomial interpolation over the unknown domain produces a good initial guess for the gradient descent minimization of WGLF that makes convergence to the steady state much faster for all choices of the forcing terms.

In most of cases the WGLF inpainting methods with the spatial vs. the edge fidelity produce almost identical results. Indeed, Fig.3.4 shows the simplest example of stripe reconnection. Having the missing area and the parameter ϵ fixed, the models with only the spatial fidelity term and only the edge fidelity give the same output when the image evolution starts from an interpolation over the missing area. Nevertheless, the dynamics of the gradient descent solution might be very different if one starts its evolution from a guess relatively far from the target steady state. An example of an inpainting problem and the corresponding choice of significant edge information can be seen in Fig.3.2, Fig.3.4. The latter figure shows that the problem set–up, in particular, the edges included in the fidelity term, depend on the choice of the wavelet function and may produce different results. As it can be seen in Fig.3.4, the use of
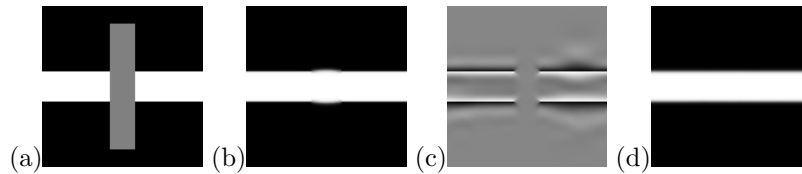

(a)   (b)   (c)   (d)

FIG. 3.4. *(a) The inpainting problem: black and white parts of the image are known, the gray area is to be inpainted, (b) minimizer of WGLF with spatial forcing term only, (c) the set of "significant" modes for the edge–preserving term of WGLF, (d) minimizer of WGLF with the edge–preserving forcing term only*

spatial fidelity results in a more exact contrast preservation, while the edge fidelity eliminates any contrast jumps near the boundary of the missing region. Connecting ability of both models is the same: the stripe is reconnected if the width of the

missing area (gray occlusion in Fig.3.4(a)) does not exceed the width of the stripe. Next subsection addresses with the question of using non-local information in the inpainting model and considers those cases when the missing area is wider than the stripe itself.

The choice between the spatial vs. edge–preserving wavelet fidelity terms is usually defined by the specifics of each image processing application. While the spatial forcing term works great for the classical inpainting problem when the unknown domain is defined a priori, the edge–preserving fidelity is of help in cases when, given a corrupted grayscale image, one needs to recover a binary one. A typical example of such situation is shown in Fig.3.5, where the task is to recover the black and white text from the scanned image of an old document (Declaration of Independence, source: *www.ushistory.org*). In order to use the spatial fidelity, one would need to allow semi–supervised processing: to obtain the user choice of the average text color and background color and then threshold the image. The edge fidelity term, on the other hand, does not require this, being defined by the adaptive coefficient thresholding described in 3.4.1. The result of the inpainting by minimizing the WGLF with the forcing term is shown in Fig.3.5.

(a)   (b)   (c)
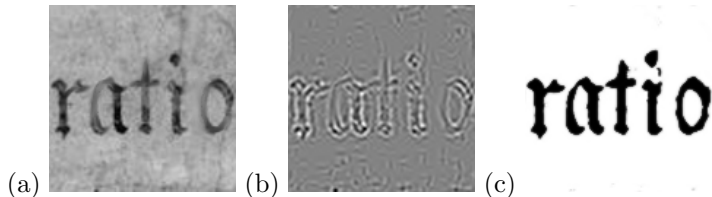
FIG. 3.5. *Advantage of the edge–preserving fidelity term for the binary text inpainting. (a) Given image (grayscale, scanned, with a lot of paper defects and color fading); (b) edges used in the forcing term of WGLF; (c) result of the WGLF inpainting with the wavelet-domain forcing term*

3.5.2. Direction–dominated inpainting. In certain cases, the dominating direction can be clearly inferred from the known part of the image. However, the energy functional we discussed before only uses information from a certain neighborhood of the missing area in the recovery of unknown intensity values. Thus, it fails to restore certain details (connections) that would seem obviously necessary to any person performing the manual image recovery. This drawback can be eliminated by introducing an additional parameter of the modified WGL energy that depends on an individual image: we allow the horizontal, vertical and diagonal components of the wavelet decomposition have variable weights in the expression for the Besov seminorm. Namely, given a function f, we replace the Besov seminorm $|f|_B^2$ in the first term of the WGL energy by the following "weighted" seminorm, which will be denoted $|f|_B^*$:

$$(|f|_B^*)^2 = a_h|f^h|_{B^2} + a_v|f^v|_{B^2} + a_d|f^d|_{B^2}, \quad a_h, a_v, a_d > 0.$$

Here f^h, f^v and f^d denote the horizontal, vertical and diagonal components of the function f. This modification to the WGL energy functional was also considered in [12] in the context of minimizing anisotropic surface tension functionals that produces the Wulff shape minimizers.

The implementation of the upgraded inpainting algorithm is completely analogous to the one described above except for an additional step that is performed before starting the gradient descent minimization. One determines relatively significant coefficients of the wavelet decomposition of the known part of a grayscale image

and concludes whether there is a direction such that respective coefficients dominate the others. If so, the WGL part of the inpainting energy functional gets updated by assigning a smaller coefficient to the dominant seminorm component, thus allowing more edges to form in the respective direction.

$$WGL^*[a](f) = \frac{\epsilon}{2}(a_h|f^h|_{B^2} + a_v|f^v|_{B^2} + a_d|f^d|_{B^2}) + \frac{1}{4\epsilon}\int f(x,y)^2(f(x,y)-1)^2 dx dy.$$

Minimizing this energy with forcing terms instead of the original WGL allows the recovered image to have a similar relation between its edge directions as the one in the original image.

Example in Fig.3.6 below shows inpainting of a horizontal stripe using the energy that prioritizes horizontal wavelet components. The sequence of images represents the gradient descent solution at consecutive time moments, the last image shows the output of the method — the recovered steady state. As one can see, the adaptively
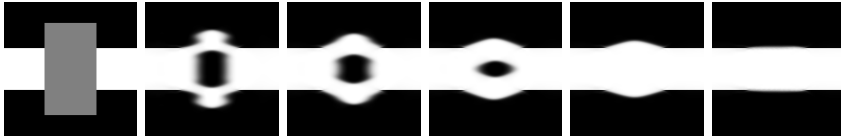
FIG. 3.6. *Solution of the gradient descent equation as time increases: the initial condition with the gap(gray) of width 50px intersecting the white stripe of width 40 px, solution after 5000, 10000, 15000, 30000, 150000 iterations.*
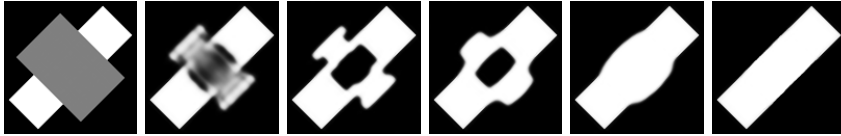
FIG. 3.7. *Solution of the gradient descent equation as time increases: the initial condition — same as in Fig.3.6, but rotated by 45°, and the solution after 250, 500, 4000, 20000, 100000 iterations.*

chosen weights assigned to the coefficients of different directions allow to achieve the contour reconnection which would otherwise be impossible in cases when the gap width exceeds the width of the stripe.

4. Modified WGL energy for denoising, segmentation, superresolution.

4.1. Inpainting approach to binary denoising. One can interpret denoising of binary images as a special type of inpainting where the entire image is treated as missing. Therefore, the inpainting model is used with $\mu_s \equiv 0$. The spatial mask, being a zero matrix, requires no locally adaptive thresholding, hence, the significant features are extracted from the image itself using absolute or relative Besov thresholding.

Denoising of grayscale and color images can be performed via bitwise denoising. The intensity matrix of each color component of an RGB image can be rewritten in the dyadic form,

$$I_{color}(x) = \sum_{j=0}^{\#bits} I_j(x)2^{-j} \quad color \in \{red, \ green, \ blue\}$$
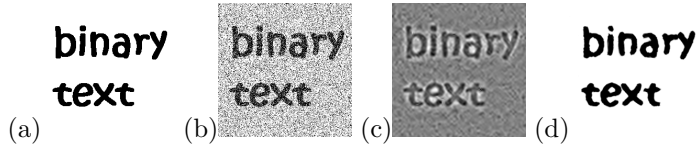
25

FIG. 4.1. *(a) Original image, (b) noisy image, SNR=7.7(dB), (c) chosen edge set, (d) output.*

and processed by denoising each dyadic digit of intensity I_j as an independent binary matrix followed by backward synthesis. The scales that can be included into the wavelet forcing term depend on the quality of data: the clearer is the image the finer scale can be enforced within denoising. However, the properties of the output are affected by the choice of the interface scale ϵ and the fidelity weight μ as well, which one can also observe in Fig.4.2.
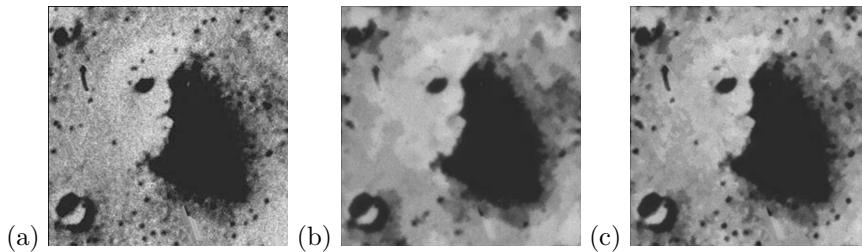
FIG. 4.2. *Grayscale image denoising (a) a noisy photo of the Mars surface taken in the 80th, 512×512 ; (b) denoised image, $\epsilon = 0.1$, $\mu = 0.05$, 6 levels of decomposition in the DB4 wavelet basis, 1 finest scale ignored (c) same setting as in (b) except for $\epsilon = 0.0625$.*

4.2. Superresolution.

4.2.1. Binary superresolution. Consider an image f represented by a matrix (f_{k_1,k_2}) of size $2^n \times 2^n$ that requires $\times 2^m$ magnification. We can identify the image matrix with a function defined on $[0,1]^2$ that belongs to the MRA subspace V_n

$$f(x) = \sum_{\vec{k}} f_{k_1,k_2} \phi_{n,\vec{k}}(x), \ x \in [0,1]^2.$$

(any image matrix is considered in this sense when one applies a discrete wavelet transform to it: we identify the spatial values and the approximation coefficients at the highest level of wavelet decomposition). Now, a matrix F of size $2^{n+m} \times 2^{n+m}$ can be obtained by sampling $f(x)$ on a refined grid of respective size: $F_{p_1,p_2} = f(\frac{p_1-1}{2^{n+m}}, \frac{p_1-1}{2^{n+m}})$. In other words, a candidate for the superresolved image matrix can be obtained by **wavelet interpolation**: representing the initial discrete signal as a finite linear combination of wavelet basis functions, and discretizing on a new, refined, grid..

The result of wavelet interpolation applied to a binary matrix f (as in Fig.4.3(a)) is not necessarily binary, moreover, it is usually grayscale — Fig.4.3(b).

Now, the problem of superresolution can be addressed as the question of denoising and edge–sharpening of the grayscale initial guess. The solution is obtained by finding minimizers to the WGLF energy with the edge preserving fidelity term only:
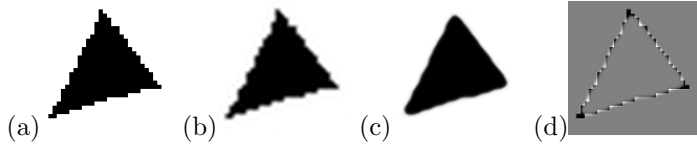
$$E(u) = WGL(u) + \frac{\lambda}{2}|P_\Lambda(u-f)|_B^2.$$

FIG. 4.3. *(a) Initial binary image of size 32x32 (a black triangle on white background), (b) the result of wavelet interpolation(using DB4) performed to increase the image 4 times, (c) gradient descent solution, $\epsilon = 0.125$, $\mu_w = 0.025$ after 500 iterations, (d) difference between the output (c) and the original image after ×4 dilation (as in 'Paint' picture editor).*

by the gradient descent method with the defined initial guess. In the absence of any additional information, index set Λ should be restricted to N coarsest scales, where N is the maximum depth of the wavelet decomposition for the original image. Here our idea is somewhat related to the wavelet pan-sharpening technique [19], where the missing fine scale information is recovered by a variational methods that preserves given data on coarser scales.

Essentially, 2^m-times magnification implies that $\log_2 m$ finest scales of the stretched image decomposition are deliberately excluded from the forcing term, as details of those scales bear no reliable information. The spatial fidelity term is identically zero. Fig.4.3 (c) shows the result of superresolution method applied to the initial image (Fig.4.3(a)) to achieve ×4 magnification. In this case two finest scales of the initial guess contain unverified information and, hence, are excluded from the edge-forcing term.

4.2.2. Grayscale and color superresolution. Superresolution of images in general can be performed bitwise, by applying a binary superresolution technique to each separate bit of a grayscale or color image. Namely, the image intensity matrix, in its dyadic representation (BIN),

$$I(x) = \sum_{j=0}^{\#bits} I_j(x)2^{-j} \qquad (BIN)$$

can be processed by "superresolving" each dyadic digit of intensity I_j as an independent binary matrix.
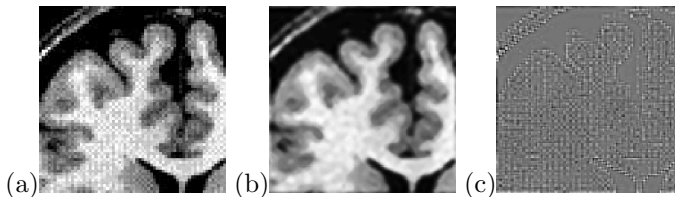

FIG. 4.4. *Example of a bitwise grayscale ×4 superresolution of a brain image: (a) the under-resolved image, (b) the result of processing and (c) their difference. The structure of the difference shows that the relevant edge information was preserved, while the pixellation effect was eliminated.*

4.3. Inpainting approach to segmentation.

4.3.1. Segmentation problem revised. Binary image segmentation, i.e. separation object/background can be also viewed as a problem of *inpainting* inside the area between the known object part and the known background part while preserving
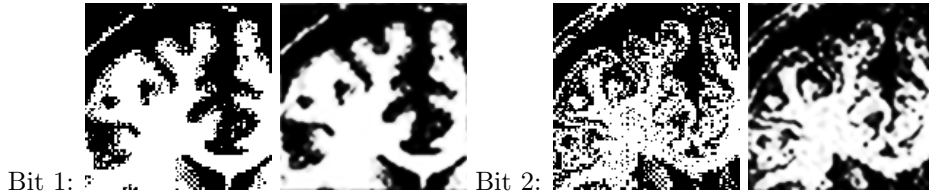
Bit 1:   Bit 2:

FIG. 4.5. *Bitwise ×4 superresolution of a brain image. First two original bits of the underresolved image (first two bits), and the output of binary superresolution.*

the edges (specifically, those that separate the latter two). The wavelet fidelity term is obtained by wavelet thresholding with zero mask, the spatial term may be either absent or localized in the preclassified (known object and known background) areas.

Fig 4.6 demonstrates this approach applied to the problem of segmenting the camera-man silhouette from the background of the image shown in Fig.4.6(a). We assume that the known data consists of two preclassified areas: those are colored in white in Fig.4.6 (b),(c) and belong to the "known background" and the "known object" respectively. The binary output of the WGLF technique, i.e. completed classification of all pixels, is shown in Fig.4.6 (f).

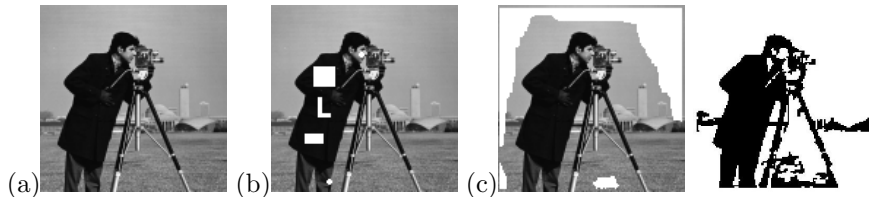(a)   (b)   (c)

FIG. 4.6. *Example of grayscale image segmentation using the binary WGLF inpainting model with the edge–preserving forcing term. (a) The original image, (b) the white areas indicate the "known" object pixels, (c) the white areas indicate the "known" background pixels, (d) the result of the WGLF segmentation.*

In cases when it is needed to extract specified objects from the given image, one can use a similar procedure with incorporated self-correction. All areas colored as "object" that do not have connection to the originally given "known object" area, should be disregarded and filled with the background color.

Fig.4.7 demonstrates an example of the aerial photo segmentation. The output image in this case is not strictly black and white, since the edge forcing term was assigned a large weight, and thus prioritized over the double–well potential. The final output can be binarized by simple thresholding of pixel values.

4.3.2. Extraction of binary contours from grayscale images. There is a class of applications which requires the concepts of inpainting and segmentation to merge. Examples of such problems include the text recovery from occluded noisy images or inpainting of the road map based on possibly occluded aerial photos.

The process of recovering specific binary contours from grayscale images follows the steps below.
   1. Analyzing the scale of the target contour (letter, road) and choosing appropriate scaling parameters.
   2. Processing (modification) of the image with the goal of tracing all edges (contours, letters, roads) while ignoring occlusions, in whichever way those
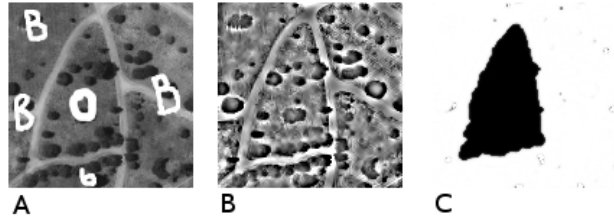
FIG. 4.7. *A. Given image (source: Google Maps) with partial classification: pixels inside the letter "O" belong to the "object", pixels within the letters "B" belong to the "background" ; B. edges used in the forcing term; C. result of the WGLF segmentation*
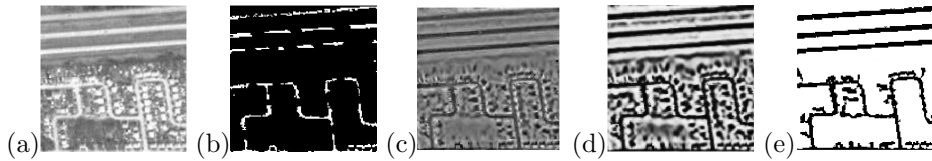

FIG. 4.8. *(a) A grayscale aerial photo (source: Google Maps); (b) partial indicator of the roads (known road pixels are white); (c) edges included in the forcing term initially; (d) intermediate state of gradient descent solution; (e) result of the self–correction: elements without connections to the "known road" areas were removed.*

are specified, until the system reaches its steady state. Provide a binary result: contours vs. background.

3. Correction of the entire contour set by excluding extra components, if there is any a priori classification of "text" or "occlusion" vs the "background" available.

Consider a problem of road-map recovery out of a grayscale aerial image and partial classification of the roads. The term "partial classification" refers to the case when the complete classification of each image pixel as "known" or "unknown" is not available, instead, one is given an incomplete indicator of road pixels (specifies the pixels that are "known to belong to roads"), with no information about the rest. An example of partial classification is shown in Fig.4.8(a),(b).

In this case, the output image is expected to be binary and represent a completed road indicator function (the updated/recovered road map). This problem setting is different from the problem of grayscale road inpainting given predefined occlusions, as it aims to recover the complete road silhouette rather than reconstruct the grayscale intensity values behind a specific occlusion.

The given image Fig.4.8(a) serves as a source for "reference edges" (obtained by non–linear thresholding), i.e. edges included in the forcing term before evolution starts Fig.4.8(c). The known road indicator can be used to construct an initial guess to speed up the algorithm convergence.

As we mentioned in the discussion of the inpainting approach to segmentation, the procedure incorporates **self–correction**. After every N iterations Fig.4.8 (d) we detect the new set of roads and "erase" connected components of this set that do not intersect with the initial road indicator as well as remove corresponding edges from the reference edge set. Such artifacts may occur due to the fast "phase separation" around near-binary values in the initial condition. The example in Fig.4.9 shows the results of processing via the same technique.
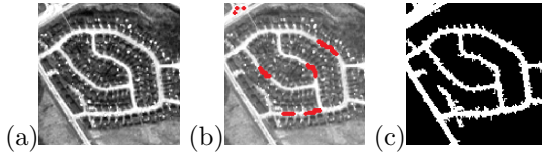
FIG. 4.9. *(a) Given grayscale image (source: Google Maps), (b) partial road classification ("known road pixels" marked as red), (c) the result of road contour detection and self-correction.*

An analogous approach can be successfully applied to the text extraction from noisy images.

5. Conclusion. We considered the problem of the WGL energy minimization by gradient descent and proved the existence of a weak solution and its convergence to a steady state which is an extremum of the WGL energy. This argument can be generalized for the gradient descent minimization associated with a modified WGL energy (completed with a spatial and an edge preserving forcing terms). The latter provides a highly adaptable variational tool for a variety of image processing problems that involve recovering and preserving important edge information. Since the described techniques involve the WGL energy that was shown to converge to a weighted TV seminorm, i.e. is anisotropic, one of the next possible steps to improving those methods is to consider a similar isotropic energy. For instance, this energy modification can be achieved by replacing the wavelet components with their analogues defined via curvelets or shearlets, which, however, involve numerical implementations of higher complexity.

REFERENCES

[1] M. Bertalmio, A. Bertozzi, and G. Sapiro. "Navier-Stokes, fluid dynamics, and image and video inpainting". *IEEE Computer Vision and Pattern Recognition (CVPR), Hawaii*, 1(I):355–362, December 2001.
[2] M. Bertalmio, G. Sapiro, C. Ballester, and V. Caselles. Image inpainting. In Kurt Akeley, editor, *Siggraph 2000, Computer Graphics Proceedings*, pages 417–424. Wesley Longman, 2000.
[3] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, 2003.
[4] A. Bertozzi, S. Esedoglu, and A. Gillette. Analysis of a two-scale Cahn-Hilliard model for image inpainting. *Multiscale Modeling and Simulation*, 6(3):913–936, 2007.
[5] H. Brezis. Monotone operators, nonlinear semigroups and applications. In *Proceedings of the International Congress of Mathematicians, Vancouver*, 1974.
[6] J.F. Cai, R.H. Chan, L.X. Shen, and Z.W. Shen. Simultaneously inpainting in image and transformed domains. *Numer. Math.*, 112(4):509–533, 2009.
[7] A. Chambolle, R. A. DeVore, N. Lee, and B. J. Lucier. Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–333, 1998.
[8] T. F. Chan, J. Shen, and H.-M. Zhou. Total variation wavelet inpainting. *Journal of Mathematical Imaging and Vision*, 25(1), July 2006.
[9] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. Lecture Notes in Statistics: Wavelets and Statistics, New York: Springer-Verlag, 1995.
[10] I. Daubechies. *Ten lectures on wavelet.* SIAM, 1992.
[11] J. Dobrosotskaya and A. Bertozzi. A Wavelet-Laplace variational technique for image deconvolution and inpainting. *"IEEE Transactions on Image Processing"*, 17(5), 2008.

[12] J. Dobrosotskaya and A. Bertozzi. Wavelet analogue of the Ginzburg-Landau energy and its Γ-convergence. *Accepted to "Interfaces and Free Boundaries"*, 2009.

[13] J. Dobrosotskaya, M. Ehler, E. King, R. Bonner, and W. Czaja. Sparse representation and variational methods in retinal image processing. In Keith E. Herold, Jafar Vossoughi, and William E. Bentley, editors, *26th Southern Biomedical Engineering Conference SBEC 2010, April 30 - May 2, 2010, College Park, Maryland, USA*, volume 32 of *IFMBE Proceedings*, pages 361–364. Springer, 2010.

[14] Bruce K. Driver. *Analysis Tools with Applications*.

[15] M. Elad, J. L. Starck, P. Querre, and D. L. Donoho. Simultaneous cartoon texture image inpaitning using morphological component analysis. *Appl. Comput. Harmon. Anal.*, 19:340–358, 2005.

[16] Lawrence C. Evans. *Partial differential equations*. AMS.

[17] S. Mallat. *Wavelet Tour of Signal Processing*. Academic Press, September 1999.

[18] Yves Meyer. *Wavelets and Operators*. Cambridge Univ. Press, 1992.

[19] M. Moeller, T. Wittman, and A. L. Bertozzi. Variational Wavelet Pan-Sharpening. *submitted to IEEE Trans. Geosci. and Remote Sensing*, 2008.

[20] I. Ya. Novikov and S. B. Stechkin. Basic wavelet theory. *Russian Mathematical Surveys*, 53:53–128, 1998.

[21] M. Welk, G. Steidl, and J. Weickert. Locally analytic schemes: A link between discussion filtering and wavelet shrinkage. *Applied and Computational Harmonic Analysis*, (24):195 – 224, 2008.