Nonlocal Anisotropic Discrete Regularization for Image, Data Filtering and Clustering

Vinh Thong Ta, Sébastien Bougleux, Abderrahim Elmoataz, Olivier Lézoray

Université de Caen Basse-Normandie, GREYC CNRS UMR 6072
6 Boulevard Maréchal Juin, F-14050 Caen Cedex France.

Abstract

In this paper, we propose a nonlocal anisotropic discrete regularization on graphs of arbitrary topologies as a framework for image, data filtering and clustering. Inspired by recent works on nonlocal regularization and on the TV digital filter, a family of discrete anisotropic functional regularization on graphs is proposed. This regularization is based on the $\mathcal{L}_p$-norm of the nonlocal gradient and the discrete p-Laplacian on graphs. It can be viewed as the discrete analogue on graphs of the continuous p-TV anisotropic functionals regularization formulations. After providing definitions and algorithms to resolve such a discrete nonlocal anisotropic regularization, we show its properties for filtering, clustering on different types of data living on different graph topologies (image, data). In particular we investigate the cases of $p = 2$, $p = 1$ and $p < 1$, this latter being very few considered in literature.

Key words: anisotropic discrete regularization; nonlocal operators; graph p-Laplacian; filtering; clustering

1 Introduction

Image regularization, based on variational and partial differential equations (PDEs) approaches, are one of the most important tools in image process-

Email addresses: vinhthong.ta@unicaen.fr (Vinh Thong Ta), sebastien.bougleux@greyc.ensicaen.fr (Sébastien Bougleux), abder.elmoataz@greyc.ensicaen.fr (Abderrahim Elmoataz), olivier.lezoray@unicaen.fr (Olivier Lézoray).
[1] This work was partially supported under a research grant of the ANR Foundation (ANR-06-MDCA-008-01/FOGRIMMI) and under a grant of FEDER/Critic3-3273 funds.

ing and computer vision. They have been applied to handle a wide variety of problems such as denoising, segmentation, optical flow or object tracking. A complete review regarding the theory and the applications of this area can be found, in [1, 5, 16, 32] and references therein. However, partial and variational methods have some limitations in the functionals used in regularization processes, such as the anisotropic diffusion models [26, 33], the total variation (TV) models [16, 29] or the active contour models [32]. Indeed, these methods are based on derivatives which only consider local features of the data.

Recently, the use of nonlocal interactions to capture the complex structure within images, has received a lot of attention. This methodology has shown to be very effective, and allows more flexibility in the regularization processes. Filters based on nonlocal interactions shown superior abilities for image and video processing such as denoising [14, 22], enhancement, demosaicing [13], texture analysis and synthesis [27]. Kindermann and Osher [23] were the first to interpret the nonlocal means filters and neighborhoods filters as nonlocal regularization functionals. Later, Guilboa and Osher [20] proposed a nonlocal functional, based on weighted differences, for image regularization and semi-supervised segmentation. These works can be regarded as the nonlocal analogues of total variation models for image regularization. Moreover, the use of nonlocal functionals, into regularization and diffusion processes not new. Indeed, it has been widely used to model diffusion processes arising from statistical mechanics [6], or neural activity modelization [34]. For recent works, see [3, 9] and references therein.

Most of the proposed nonlocal regularization functionals have been formulated for image processing, where images are expressed as continuous functions on continuous domains. Then, a continuous energy functional is considered and classically solved by the corresponding continuous Euler-Lagrange equations or its associated flows. Unfortunately, the discretization of these equations is difficult for high dimensional data, or functions defined on this data. It is also the case for irregular domains. Moreover, in many research fields such as geography, data-mining, social sciences, or robotics, data can be represented by graphs. The interactions within the data can be expressed by functionals on graphs, and the related problems can be treated in terms of regularization.

To regularize images or discrete data on irregular domains, another methodology consists in performing complete digitization of continuous models [25]. The latter approach directly considers discrete variational problems, and therefore works on general discrete data. An example of such a digitization is the TV digital filter, introduced and analyzed by [15] in the context of image restoration. This filter can be interpreted as a precise translation of the TV model [29] in discrete settings on unweighted graphs. Inspired by this work, we have recently proposed a general p-TV discrete model on weighted graphs [11, 19]. A similar approach has also been investigated in the context of semi-supervised learning on graphs [37].

Our previously proposed regularization is based on the $\mathcal{L}_2$-norm of the nonlocal gradient and on the discrete p-Laplacian on graphs. It can be viewed as the discrete analogue of the continuous p-TV isotropic regularization. We have shown that our discrete p-TV isotropic regularization framework leads to a family of linear and nonlinear processing on graphs. In particular, this family of methods includes exact expression of several local and nonlocal filters such as the bilateral filter [30, 31], the TV digital filter, or the nonlocal means filter [14]. Moreover, for image processing, we have also shown that the continuous local and nonlocal regularizations, based on p-TV models, are the continuous analogue of our discrete regularization approach. Indirectly and naturally, our methodology provides a discrete extension of these continuous regularization methods for any discrete data or functions.

In this paper, we propose to extend our isotropic regularization framework by introduce a new family of discrete anisotropic regularization functionals. This new family is based the $\mathcal{L}_p$-norm, and generalize significantly our previous studies. Moreover, it can be considered as the nonlocal discrete analogue of the p-TV anisotropic regularization in the continuous case [21].

Let $G = (V, E, w)$ be a weighted graph which consists in a set of vertices V, a set of edges $E \subset V \times V$, and a weight function w defined on the edges. The discrete p-TV anisotropic regularization of a function $f^0 : V \to \mathbb{R}$ is performed by the following energy minimization:

$$\min_{f:V\to\mathbb{R}} \left\{ E_w(f, f^0, \lambda, p) = R_w(f, p) + \tfrac{\lambda}{2}\|f - f^0\|_2^2 \right\}, \tag{1}$$

where $p \in [0, +\infty[$ is the smoothness degree, and $\lambda \geq 0$ is the Lagrange multiplier or fidelity parameter. The first term in (1), is defined as $\frac{1}{2p}\sum_{v\in V}|\nabla_w f(v)|_p^p$, where $|.|_p$ is the $\mathcal{L}_p$ norm and $\nabla_w f$ is the weighted gradient of f over G. The second term is the fidelity term and it estimates the variations between the functions f^0 and f. For the case of $p = 2$, one can easily show that Eq. (1) corresponds to the Tikhonov regularization. For several values of p, this family of functionals is related to numerous processing methods involve in image restoration, graphs cuts or random walks techniques [2].

This paper is organized as follows. Firstly we introduce the basic operators needed for the formulation, and for the resolution, of the proposed minimization problem. Then, we propose two different diffusion processes to approximate its solution. Finally, the obtained regularization is applied to filter and to classify images and discrete data.

2 Discrete operators on weighted graphs

This section recall useful notations on weighted graphs and define the basic discrete operators involved in the proposed regularization framework. Similar definitions and properties have also been used in the context of spectral graph theory [17], differential calculus on graphs [8,24], semi-supervised learning [35,36], or image and mesh processing [10,11,15,25]. In these works, the discrete operators are usually defined in order to formalize and to solve isotropic regularizations, which leads to diffusion processes based on isotropic graph Laplacians or p-Laplace operators. In the following, we propose to construct analogue operators for the anisotropic framework. In particular, we define the discrete anisotropic p-Laplace operator.

2.1 Preliminary definitions

A weighted graph $G = (V, E, w)$ is composed of a finite set V of N vertices, a finite set $E \subset V \times V$ of edges, and a weight function $w : E \to \mathbb{R}^+$. An edge of E, which connects two vertices u and v of V, is noted uv. In this paper, graphs are supposed to be connected, with no self-loops or multiple edges, and undirected (for each edge $uv \in E$, $vu \in E$). This implies that the weight function w is symmetric, i.e. $w_{uv} = w_{vu}$ for all edge $uv \in E$. The weight function measures the dissimilarity of two vertices of the graph. When $w_{uv} \to 0$, the two vertices u and v are dissimilar. By convention, we set $w_{uv} = 0$ if the vertices are not connected by an edge of E.

Let $\mathcal{H}(V)$ be the Hilbert space of real-valued functions on the vertices of a graph $G = (V, E, w)$, with $V = \{v_1, \ldots, v_N\}$. Each function $f \in \mathcal{H}(V)$, $f : V \to \mathbb{R}$, assigns a vector $f(v_i)$ to each vertex $v_i \in V$. The function f forms a finite N-dimensional space. It can be thought as a column vector $f = [f(v_1), \ldots, f(v_N)]^T$. By analogy with functional analysis in continuous spaces, the integration of f over the graph, is noted $\int_V f = \sum_V f$. Similarly, let $\mathcal{H}(E)$ be the Hilbert space of real-valued functions $F : E \to \mathbb{R}$, defined on the edges of the graph. These two spaces are endowed with the usual inner products:

$$\langle f, g \rangle_{\mathcal{H}(V)} = \sum_{u \in V} f(u)g(u), \quad \langle F, G \rangle_{\mathcal{H}(E)} = \sum_{u \in V} \sum_{v \sim u} F(uv)G(uv),$$

where $f, g : V \to \mathbb{R}$ and $F, G : E \to \mathbb{R}$.

The *difference operator* of a function $f : V \to \mathbb{R}$, noted $d : \mathcal{H}(V) \to \mathcal{H}(E)$, is defined on an edge $uv \in E$ by:

$$(df)(uv) = \frac{f(v) - f(u)}{\mu_w(u,v)},$$

where $\mu : V \times V \to \mathbb{R}^+$ is a similarity measure which depends on the weight function w. In the sequel, we restrict ourselves to $\mu_w(u,v) = 1/\sqrt{w_{uv}}$. This implies the following definition of the difference operator:

$$(df)(uv) = \sqrt{w_{uv}}(f(v) - f(u)) \tag{2}$$

The *directional derivative* of the function f, at a vertex u, and over an edge uv, is defined to be:

$$\partial_v f(u) = (df)(uv) = \sqrt{w_{uv}}(f(v) - f(u)). \tag{3}$$

The difference operator and the directional derivative are antisymmetric, i.e. $(df)(uv) = -(df)(vu)$. They also share the following property with the continuous definition of the derivative of function defined in the Euclidean space:

$$f(u) = f(v) \Rightarrow (df)(uv) = 0. \tag{4}$$

The *adjoint operator* of the difference operator, noted $d^* : \mathcal{H}(E) \to \mathcal{H}(V)$, is defined by:

$$\langle df, H \rangle_{\mathcal{H}(E)} = \langle f, d^* H \rangle_{\mathcal{H}(V)}, \qquad \forall f \in \mathcal{H}(V), H \in \mathcal{H}(E). \tag{5}$$

Then, from the expressions of the inner products in $\mathcal{H}(E)$ and $\mathcal{H}(V)$, one can deduce the expression of the adjoint operator at a vertex $u \in V$:

$$(d^* H)(u) = \sum_{v \in V} \sqrt{w_{uv}}(H(vu) - H(uv)). \tag{6}$$

The adjoint is a linear operator which measures the flow of H over the graph. By analogy with continuous differential operators, the *divergence operator* of the function H is defined as $div\, H = -d^* H$. Then, one can show that any function $H \in \mathcal{H}(E)$ has a null flow (divergence theorem):

$$\sum_{u \in V} (div\, H)(u) = \sum_{u \in V} \sum_{v \in V} \sqrt{w_{uv}}(H(uv) - H(vu)) = 0. \tag{7}$$

Remark. *There exists other expressions of the difference operator, depending on the context. For example, the difference operator defined in [35, 36] is formulated as:*

$$(df)(uv) = \sqrt{\frac{w_{uv}}{\delta_v}} f(v) - \sqrt{\frac{w_{uv}}{\delta_u}} f(u),$$

where $\delta_u = \sum_{v\sim u} w_{uv}$ is the degree of the vertex u in the graph. One can note that this formulation does not respect the property (4). Moreover, its adjoint operator does not satisfy the divergence theorem (Eq. (7)).

2.3 Gradient operator and local variations

The *gradient operator* of a function $f : V \to \mathbb{R}$, at a vertex $u \in V$, is the vector operator defined by:

$$\nabla_w f(u) = \Big(\partial_v f(u) \,:\, v \sim u\Big)^T = \Big(\partial_{v_1} f(u), \ldots, \partial_{v_k} f(u)\Big)^T,$$

where $v_1, \ldots, v_k$ are the neighbors of the vertex u.

The *local variation* of the function f measures its regularity in the neighborhood of a vertex. Several norms of $\nabla_w f(u)$ can be used. In this paper, we propose to use the $\mathcal{L}_2$-norm:

$$|\nabla_w f(u)|_2 = \left(\sum_{v\sim u} (\partial_v f(u))^2\right)^{\frac{1}{2}} = \left(\sum_{v\sim u} w_{uv}(f(v) - f(u))^2\right)^{\frac{1}{2}}, \qquad (8)$$

and more generally the $\mathcal{L}_p$-norm:

$$|\nabla_w f(u)|_p = \left(\sum_{v\sim u} |\partial_v f(u)|^p\right)^{\frac{1}{p}} = \left(\sum_{v\sim u} w_{uv}^{\frac{p}{2}} |f(v) - f(u)|^p\right)^{\frac{1}{p}}. \qquad (9)$$

The $\mathcal{L}_p$-norm of the local variation is a seminorm for $p \geq 1$, and it is not a norm for $p < 1$.

When the set of vertices V represents a set of organized data, such as digital images or meshes, the initial organization, which is a graph by nature, traduces local interactions between the data. On this initial graph, the gradient operator involves these local interactions. A gradient operator, which involves nonlocal interactions, can be obtained by constructing neighborhood graphs from the initial organization. One can remark that in both cases, the expression of the local variation is the same. However, due to the fact that $w_{uv} = 0$ if $uv \notin E$, a totally nonlocal expression of the local variation can be written as :

$$| \nabla_w^{NL} f(u)|_p = \left(\sum_{v\in V} |\partial_v f(u)|^p\right)^{\frac{1}{p}} = \left(\sum_{v\in V} w_{uv}^{\frac{p}{2}} |f(v) - f(u)|^p\right)^{\frac{1}{p}}. \qquad (10)$$

The *anisotropic p-Laplace operator* $\Delta_w^p : \mathcal{H}(V) \to \mathcal{H}(V)$, of a function $f : V \to \mathbb{R}$, is defined using the difference operator:

$$\Delta_w^p f = \frac{1}{2} d^* \left(|df|^{p-2} df \right) \tag{11}$$

Its expression, at a vertex $u \in V$, is given by:

$$\begin{aligned} \Delta_w^p f(u) &= \frac{1}{2} \sum_{v \sim u} \sqrt{w_{uv}} \left(|(df)(vu)|^{p-2} (df)(vu) - |(df)(uv)|^{p-2} (df)(uv) \right) \\ &= \sum_{v \sim u} \sqrt{w_{uv}} \left(|(df)(vu)|^{p-2} (df)(vu) \right) \\ &= \sum_{v \sim u} w_{uv}^{\frac{p}{2}} |f(u) - f(v)|^{p-2} (f(u) - f(v)). \end{aligned} \tag{12}$$

Two specific cases of the anisotropic p-Laplace operator are given for $p = 1$ and $p = 2$. When $p = 1$, an interesting rewriting of Eq. (12) is given by:

$$\Delta_w^1 f(u) = \sum_{v \sim u} \sqrt{w_{uv}} \, sign(f(u) - f(v)). \tag{13}$$

When $p = 2$, $\Delta_w^2 f = \Delta_w f = \frac{1}{2} d^*(df)$, which the definition of the combinatorial (isotropic) graph Laplacian. In this case, Eq. (12) becomes:

$$\Delta_w f(u) = \sum_{v \sim u} w_{uv} (f(u) - f(v)). \tag{14}$$

In order to avoid numerical instabilities when $p < 2$, the anisotropic p-Laplace operator is regularized as:

$$\Delta_w^{p,\epsilon} f(u) = \sum_{v \sim u} w_{uv}^{\frac{p}{2}} \left(|f(u) - f(v)| + \epsilon \right)^{p-2} (f(u) - f(v)), \tag{15}$$

where $\epsilon > 0$ is fixed constant.

Remark. *There exists much more discrete expressions of the isotropic p-Laplace operator, defined as* $\Delta_w^p f = \frac{1}{2} d^* \left(| \nabla_w f|_2^{p-2} df \right)$*. Using the difference operator and its adjoint of Section 2.2, the expression of the isotropic p-Laplace operator is given by:*

$$\Delta_w^p f(u) = \sum_{v \sim u} w_{uv} (| \nabla_w f(u)|_2^{p-2} + | \nabla_w f(v)|_2^{p-2})(f(u) - f(v)).$$

See [11] for more details.

3 Discrete anisotropic regularization framework

Let $G = (V, E, w)$ be a weighted graph, and let $f^0 : V \to \mathbb{R}$ be a function of $\mathcal{H}(V)$. To regularize f^0, we propose to consider the following variational problem:

$$\min_{f \in \mathcal{H}(V)} \left\{ E_w(f, f^0, \lambda, p) = R_w(f, p) + \tfrac{\lambda}{2} \|f - f^0\|_2^2 \right\}, \tag{16}$$

where $R_w(f, p)$ is the anisotropic p-TV functional defined as:

$$\begin{aligned} R_w(f, p) &= \tfrac{1}{2p} \sum_{u \in V} |\nabla_w f(u)|_p^p, \qquad p \in]0, +\infty[\\ &= \tfrac{1}{2p} \sum_{u \in V} \sum_{v \sim u} w_{uv}^{\frac{p}{2}} |f(v) - f(u)|^p. \end{aligned} \tag{17}$$

For $p \geq 1$, both functionals in the minimizer (16) are convex, and the solution of the minimization is unique. This is not the case when $p < 1$, for which the regularization functional $R_w(f, p)$ is not convex, and the uniqueness of the minimization solution is not insured. However, this case is also considered in the following, in order to analyze the behavior of the associated diffusion processes beyond the usual bound $p = 1$, which have been very few investigated by researchers (see Section 5 and Section 6).

In particular, when $p = 1$, the regularization functional is the anisotropic total variation of a function $f : V \to \mathbb{R}$:

$$R_w(f, 1) = \tfrac{1}{2} \sum_{u \in V} |\nabla_w f(u)|_1 = \tfrac{1}{2} \sum_{u \in V} \sum_{v \sim u} \sqrt{w_{uv}} |f(v) - f(u)|. \tag{18}$$

When $p = 2$, $R_w(f, 2)$ is the classical (isotropic) regularization functional:

$$R_w(f, 2) = \tfrac{1}{4} \sum_{u \in V} |\nabla_w f(u)|_2^2 = \tfrac{1}{4} \sum_{u \in V} \sum_{v \sim u} w_{uv} (f(v) - f(u))^2, \tag{19}$$

and the minimization of $E_w(f, f^0, \lambda, 2)$ is the Tikhonov regularization of the function f^0 on a weighted graph.

Remark. *The anisotropic regularization functional $R_w(f, p)$ is the discrete weighted transcription of the continuous regularization functional of a real-valued function $f : \Omega \subset \mathbb{R}^m \to \mathbb{R}$, defined on a bounded domain Ω of the m-dimensional Euclidean space:*

$$J(f, p) = \int_\Omega | \bigtriangledown f|_p^p dx.$$

One can note that the resolution of the discrete minimization problem does not require specific boundary conditions. They are naturally encoded by the graph structure.

To establish the solution of the minimization problem (16), on can solve the following system of equations:

$$\frac{\partial}{\partial f}E_w(f,f^0,\lambda,p)\bigg|_{f(u)} = \frac{\partial}{\partial f}R_w(f,p)\bigg|_{f(u)} + \lambda((f(u)-f^0(u)) = 0, \quad \forall u \in V.$$

Then, if this system has a solution, it is the unique solution of the minimizer (16), for $p \geq 1$.

Property 1 $\frac{\partial}{\partial f}R_w(f,p)\Big|_{f(u)} = \Delta_w^p f(u)$.

PROOF. Let u_1 be a vertex of V. The $u_1{}^{\text{th}}$ term of the partial derivative of $R_w(f,p)$ is given by:

$$\frac{\partial}{\partial f}R_w(f,p)\bigg|_{f(u_1)} \overset{(17)}{=} \frac{\partial}{\partial f}\left(\frac{1}{2p}\sum_{u\in V}\sum_{v\sim u} w_{uv}^{\frac{p}{2}}|f(v)-f(u)|^p\right)\bigg|_{f(u_1)}.$$

The partial derivative depends only on the edges incident to u_1 . Let $v_1,\ldots,v_k$ be the vertices of V connected to u_1 by an edge of E. Then, using the chain rule, we have:

$$\begin{aligned}
2\,\frac{\partial R_w(f,p)}{\partial f}\bigg|_{f(u_1)} &= -\sum_{v\sim u_1} w_{u_1 v}^{\frac{p}{2}}\left(f(v)-f(u_1)\right)|f(v)-f(u_1)|^{p-2} \\
&\quad + w_{u_1 v_1}^{\frac{p}{2}}(f(u_1)-f(v_1))|f(u_1)-f(v_1)|^{p-2} \\
&\quad + \ldots + w_{u_1 v_k}^{\frac{p}{2}}(f(u_1)-f(v_k))|f(u_1)-f(v_k)|^{p-2} \\
&= 2\sum_{v\sim u_1} w_{u_1 v}^{\frac{p}{2}}(f(u_1)-f(v))|f(u_1)-f(v)|^{p-2} \\
&\overset{(12)}{=} 2\Delta_w^p f(u_1). \ \square
\end{aligned}$$

Based on Property 1, the solution becomes:

$$\Delta_w^p f(u) + \lambda(f(u)-f^0(u)) = 0, \quad \forall u \in V. \tag{20}$$

This latter equation can be considered as the discrete Euler-Lagrange equation associated to the minimization problem (16). Contrary to the continuous case, it does not involve any PDEs. By replacing the expression of the anisotropic p-Laplace operator Δ_w^p in the system (20), we obtain the following nonlinear system:

$$\sum_{v\sim u} w_{uv}^{\frac{p}{2}}|f(u)-f(v)|^{p-2}(f(u)-f(v)) + \lambda(f(u)-f^0(u)) = 0, \quad \forall u \in V. \tag{21}$$

Several methods can be used to solve this system. In particular, for $p = 1$, commonly used methods are based on graph cut techniques [12, 18]. In the sequel, we propose to use two simple methods to approximate the solution of the system (21). These methods leads to semi-discrete and discrete diffusion processes.

Semi-discrete diffusion process. The first method, considered in this paper, is the infinitesimal steepest descent method:

$$\begin{cases} f^{(0)} = f^0 \\ \dfrac{d}{dt} f^{(t)}(u) = -\Delta_w^p f(u) + \lambda(f^{(0)}(u) - f(u)), \quad \forall u \in V, \end{cases} \tag{22}$$

where $f^{(t)}$ is the parametrization of the function f by an artificial time. This is a system of ordinary differential equations. Contrary to PDEs methods, no space discretization is necessary. Its solution can be efficiently approximated by local iterative methods, such as the Euler method, given by:

$$\begin{cases} f^{(0)} = f^0 \\ \Delta = \sum_{v \sim u} w_{uv}^{p/2} (|f^{(t)}(u) - f^{(t)}(v)| + \epsilon)^{p-2} (f^{(t)}(v) - f^{(t)}(u)) \\ f^{(t+1)}(u) = f^{(t)}(u) + \tau(\Delta + \lambda(f^0(u) - f^{(t)}(u))) \end{cases} \tag{23}$$

In order to satisfy the minimization problem, the stopping criterion, of the above diffusion process, is a convergence condition. Meanwhile a fixed total number of iterations corresponds to an anisotropic semi-discrete diffusion process. When $p = 1$, an interesting expression of an iteration of the process (23) is given by:

$$f^{(t+1)}(u) = f^{(t)} + \tau \left(\sum_{v \sim u} \sqrt{w_{uv}} sign(f^{(t)}(v) - f^{(t)}(u)) + \lambda(f^0(u) - f^{(t)}(u)) \right).$$

This allows to avoid the use of the regularized anisotropic 1-Laplace operator $\Delta_w^{p,\epsilon}$. When $p = 2$, the diffusion process is the classical linear isotropic geometric diffusion based on the combinatorial Laplace operator [17]. In this case, an iteration of the process (23) is rewritten as:

$$f^{(t+1)}(u) = f^{(t)} + \tau \left(\sum_{v \sim u} w_{uv} (f^{(t)}(v) - f^{(t)}(u)) + \lambda(f^0(u) - f^{(t)}(u)) \right).$$

The behavior of diffusion process (23) is illustrated, for several values of $p \leq 2$ and graph structures (topologies and weight functions), in the context of image filtering and simplification (see Section 5).

Discrete diffusion process. To approximate the solution of the minimization problem (16), another method is to linearize the system of equations (21), which can be rewritten as:

$$(21) \Leftrightarrow f(u) = \frac{\lambda f^0(u) + \sum_{v \sim u} w_{uv}^{\frac{p}{2}} (|f(u) - f(v)| + \epsilon)^{p-2} f(v)}{\lambda + \sum_{v \sim u} w_{uv}^{\frac{p}{2}} (|f(u) - f(v)| + \epsilon)^{p-2}}, \quad \forall u \in V.$$

Using the linearized Gauss-Jacobi method, the solution of this system is summarized by the following iterative algorithm:

$$\begin{cases} f^{(0)} = f^0 \\ \gamma_{uv}(f^{(t)}) = w_{uv}^{\frac{p}{2}} (|f^{(t)}(u) - f^{(t)}(v)| + \epsilon)^{p-2} \\ f^{(t+1)} = \dfrac{\lambda f^0(u) + \sum_{v \sim u} \gamma_{uv}(f^{(t)}) f^{(t)}(v)}{\lambda + \sum_{v \sim u} \gamma_{uv}(f^{(t)})}, \quad \forall u \in V. \end{cases} \tag{24}$$

Let φ be the function defined by:

$$\varphi_{uv}(f) = \frac{\gamma_{uv}(f)}{\lambda + \sum_{v \sim u} \gamma_{uv}(f)}, \text{ if } u \neq v, \quad \text{or } \varphi_{uu}(f) = \frac{\lambda}{\lambda + \sum_{v \sim u} \gamma_{uv}(f)}.$$

Then, the diffusion process (24) is rewritten as:

$$\begin{cases} f^{(0)} = f^0 \\ f^{(t+1)}(u) = \varphi_{uu}(f^{(t)}) f^0(u) + \sum_{v \sim u} \varphi_{uv}(f^{(t)}) f^{(t)}(v). \end{cases} \tag{25}$$

At each iteration of this adaptive process, the new value $f^{(t+1)}$, at a vertex u, depends on two quantities: the original value $f^0(u)$, and a weighted average of the existing values in a neighborhood of u. Since $\varphi_{uu}(f) + \sum_{v \sim u} \varphi_{uv}(f) = 1$, this diffusion process is a forced low-pass filter. Through the values of the regularization parameter p, it describes a family of iterative filters. In particular, when $p = 2$, $\gamma_{uv}(f)$ reduces to w_{uv}, and the above discrete diffusion process corresponds to the one proposed in [10,11,25], in the context of isotropic regularization. The behavior of the diffusion process (24) is illustrated in Section 6, for several values of p and graph structures, in the context of semi-supervised image segmentation.

4 Nonlocal regularization of discrete data and related works

The regularization framework, proposed in the previous section, can be used to process any set of discrete data or function defined on these data. The data can be organized or not. Let $V = \{v_1, \ldots, v_N\}$ be a finite set of data, such that each data v_i is a vector in a given metric space. To regularize V, or a function $f^0 : V \to \mathbb{R}$, the first step is to construct a graph $G = (V, E, w)$. The edges of E can express different kind of interactions. A particular case consists in generating the fully connected graph, where $E = V \times V \setminus \{uu,\ u \in V\}$. In this case, the anisotropic regularization functional $R_w(f,p)$ (Eq. (17)) takes into account nonlocal interactions:

$$R_w(f,p) = \frac{1}{2p} \sum_{v \in V} \sum_{v \in V} w_{uv}^{\frac{p}{2}} |f(v) - f(u)|^p. \tag{26}$$

One can remark that this last expression is always true, whatever the structure of the graph is. This due to the fact that, by definition, $w_{uv} = 0$ if the vertex u is not connected to the vertex v by an edge of E. This shows that local and nonlocal regularizations have exactly the same expression when they are performed on graphs.

The nonlocal functional (26) is the discrete analogue of the continuous nonlocal functional of functions $f : \Omega \subset \mathbb{R}^m \to \mathbb{R}$, defined in a bounded domain Ω of the Euclidean space:

$$J_w(f,p) = \frac{1}{2p} \int_{\Omega \times \Omega} w_{xy}^{\frac{p}{2}} |f(y) - f(x)|^p dydx.$$

When $p = 1$, Eq. (26) is the discrete analogue of the nonlocal anisotropic functional based on differences proposed by Gilboa and Osher [21]:

$$J_w(f,1) = \frac{1}{2} \int_{\Omega \times \Omega} \sqrt{w_{xy}} |f(y) - f(x)| dydx.$$

When $p = 2$, Eq. (26) is the discrete analogue of the nonlocal isotropic functional proposed in [20, 23]:

$$J_w(f,2) = \frac{1}{4} \int_{\Omega \times \Omega} w_{xy} (f(y) - f(x))^2 dydx.$$

This last functional is the variational interpretation of a family of neighborhood filters, widely used in image processing, such as the nonlocal means filter [14].

One can remark that one iteration of the discrete diffusion process (24), for $p = 2$, $\lambda = 0$ and specific weight functions, corresponds to well-known filters. In particular, we can retrieve the bilateral filter [30,31] using the following weight function:

$$w_{uv} = \exp\left(-\frac{\|u - v\|^2}{\sigma_V^2}\right) \exp\left(-\frac{\|f^0(v) - f^0(u)\|^2}{\sigma_{\mathcal{H}(V)}^2}\right), \tag{27}$$

where σ_V and $\sigma_{\mathcal{H}(V)}$ are the variances respectively in the discrete domain V and in the space of functions $\mathcal{H}(V)$. Similarly, the expression of the nonlocal means filter [14] is given using the weight function:

$$w_{uv} = \exp\left(-\frac{\rho(F_{f^0}(v), F_{f^0}(u))}{h^2}\right), \tag{28}$$

where $F_{f^0}(u)$ is a feature vector associated to the vertex $u \in V$, which is generally a patch of vertices, centered at u. The function ρ measures the distance between the two feature vectors. With these two weight functions, the anisotropic regularization framework behaves like an iterated bilateral filter or an iterated nonlocal means filter, without updating the weights at each iteration of the diffusion processes.

When $\lambda = 0$, and for $p = 1$ and $p = 2$, our regularization functional corresponds to the functional used in spectral graph analysis, recently proposed in the context of semi-supervised image segmentation. When $p = 1$, the minimization problem is solved with graph cut techniques, and when $p = 2$ with random walks approaches [2, 17].

5 Image filtering and simplification

In this section, the behavior of the proposed anisotropic regularization framework is illustrated in the context of image filtering and simplification, for $p \leq 2$ and several graph topologies. An image $I : \Omega \subset \mathbb{Z}^2 \rightarrow \mathbb{R}$ is modelized by a neighborhood graph $G = (V, E, w)$ and a function $f^0 : V \rightarrow \mathbb{R}$. To each pixel (i, j) of the discrete domain Ω corresponds a vertex $u = (i, j)$ of V. The set E of edges is generated using the nearest neighbors of the vertices, based on the Chebyshev distance. Let $u = (i, j)$ be a vertex of V, its neighborhood is defined by:

$$\mathcal{N}_k(u) = \{v = (i', j') \in V \setminus \{u\} \ : \ \max\{|i - i'|, |j - j'|\} \leq k, \ k > 0\}.$$

Then the edge uv is in G_k if $v \in \mathcal{N}_k(u)$ (and reciprocally). G_1 is the 8-adjacency graph, and the fully connected graph is noted G_∞.

To regularize the image I (the function f^0), we use the semi-discrete diffusion process. In our experiments, we found that using $p \leq 1$, with a local or a nonlocal representation of the image, helps to preserve sharp edges during the regularization. This behavior is illustrated in Fig. 1 on an unweighted graph of 8-adjacency (G_1). The results are given for two different values of λ. One can observe the difference between the isotropic regularization with $p = 2$, and the anisotropic regularization with $p \leq 1$. The anisotropic case behaves like a simplification procedure, in which the image discontinuities are more well-preserved. This is also the case when the graph is weighted (see Fig. 2), or when it is generated from larger neighborhoods (see Fig. 3). In the nonlocal scheme, the main textures are also preserved. The local case, which can be computed efficiently, may be used in simplification and segmentation processes.

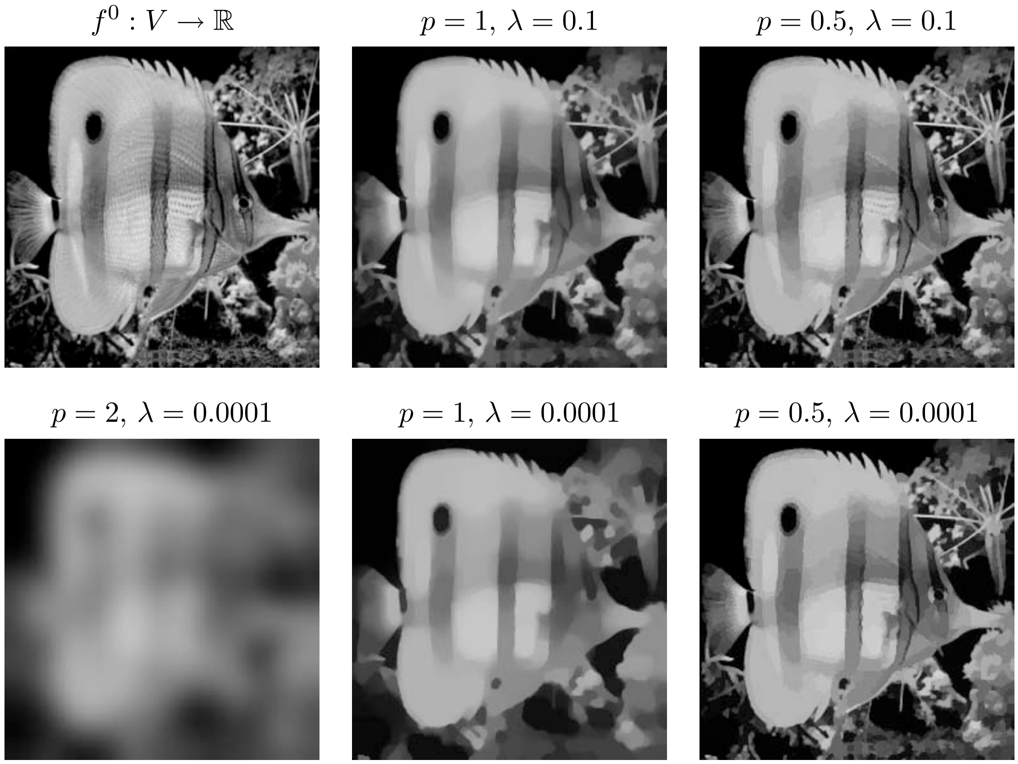


Fig. 1. Regularization of an image, on a 8-adjacency unweighted graph (G_1).

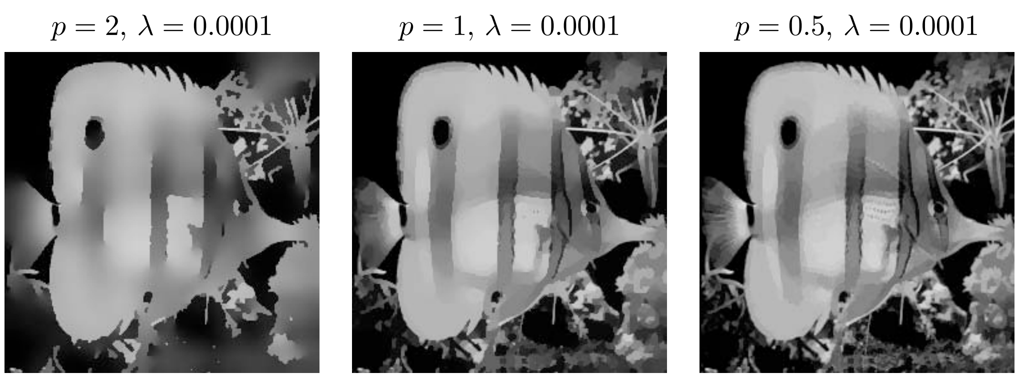


Fig. 2. Regularization of an image, on a 8-adjacency graph (G_1), weighted by the function of Eq. (27).

6 Semi-supervised classification

Recent approaches based on diffusion processes show the efficiency of label propagation to achieve manifold semi-supervised classification as in [7, 36], which use the p-Laplacian diffusion regularization. More recently [2] uses approaches based on the $\mathcal{L}_p$-norm to solve image segmentation problem. In this section, we use our discrete anisotropic regularization framework to address this learning problem for image semi-supervised segmentation and data clus-

$p = 2,\ \lambda = 0.0001$ $p = 1,\ \lambda = 0.0001$ $p = 0.5,\ \lambda = 0.0001$

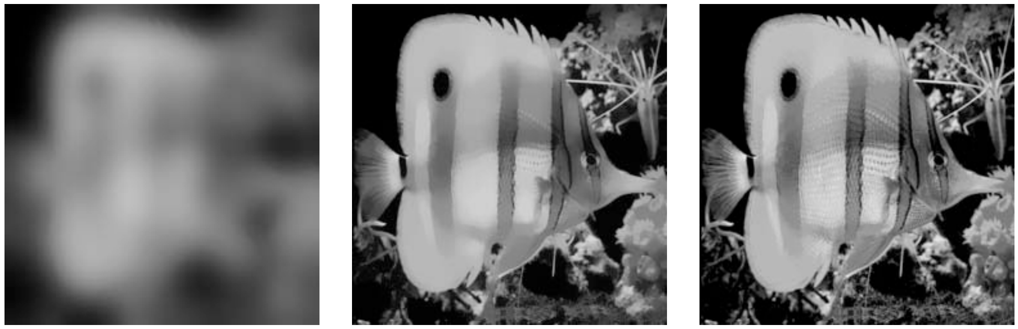


Fig. 3. Regularization of an image, on a the graph (G_4), weighted by the function of Eq. (28), with a patch of size 5×5 as a feature vector.

tering.

6.1 Problem formulation

Let $V = \{v_1, \ldots v_N\}$ be a finite set of data, where each data v_i is a vector of $\mathbb{R}^m$. Let $G = (V, E, w)$ be a weighted graph such that data are connected by an edge of E. The semi-supervised classification of the set V consists in classify the set V into k classes where the number of k classes is given. For this, the set V is composed of labeled and unlabeled data. The objective is to estimate the unlabeled data from labeled ones.

Let c_i be the set of vertices which belong to the i^{th} class. The set $C = \{c_i\}_{i=1,\ldots,k}$ is the initial set of labeled data, and the initial unlabeled data belong to the set $V \backslash C$. This is equivalent to consider k label functions $f_i^0 : V \to \mathbb{R}$ such as

$$f_i^0(v) = \begin{cases} +1 & \text{if } v \in c_i \text{ with } i \in [1, k],\ \forall c \in C \\ -1 & \text{otherwise} \\ 0 & v \in V \setminus C \end{cases} \tag{29}$$

where each f_i^0, with $i = 1, \ldots, k$, corresponds to a given class. Starting from the labeled data (the f_i^0's), the classification is accomplished by k regularization processes by estimating the resultant function $f_i : V \to \mathbb{R}$.

Using our proposed anisotropic regularization framework, this is formalized as follows:

$$\min_{f_i \in \mathcal{H}(V)} \left\{ R_w(f_i, p) + \tfrac{\lambda}{2} \|f_i(v) - g_i(v)\|_2^2 \right\}, \forall i = 1, \ldots, k.$$

We use the discrete diffusion process (24) to compute each minimization. At the end of the label propagation algorithm, one can estimate the class membership probability, for all $i \in 1, \ldots, k$, by:

$$p\Big(\omega = c_i | f(v)\Big) = \frac{f_i(v)}{\sum\limits_i f_i(v)},$$

and assign to the vertex v the most plausible one by

$$c(v) = \arg\max_i p\Big(\omega = c_i | f(v)\Big). \tag{30}$$

The following section shows the application of this classification problem to image segmentation and data clustering.

6.2 Semi-supervised image segmentation and data classification

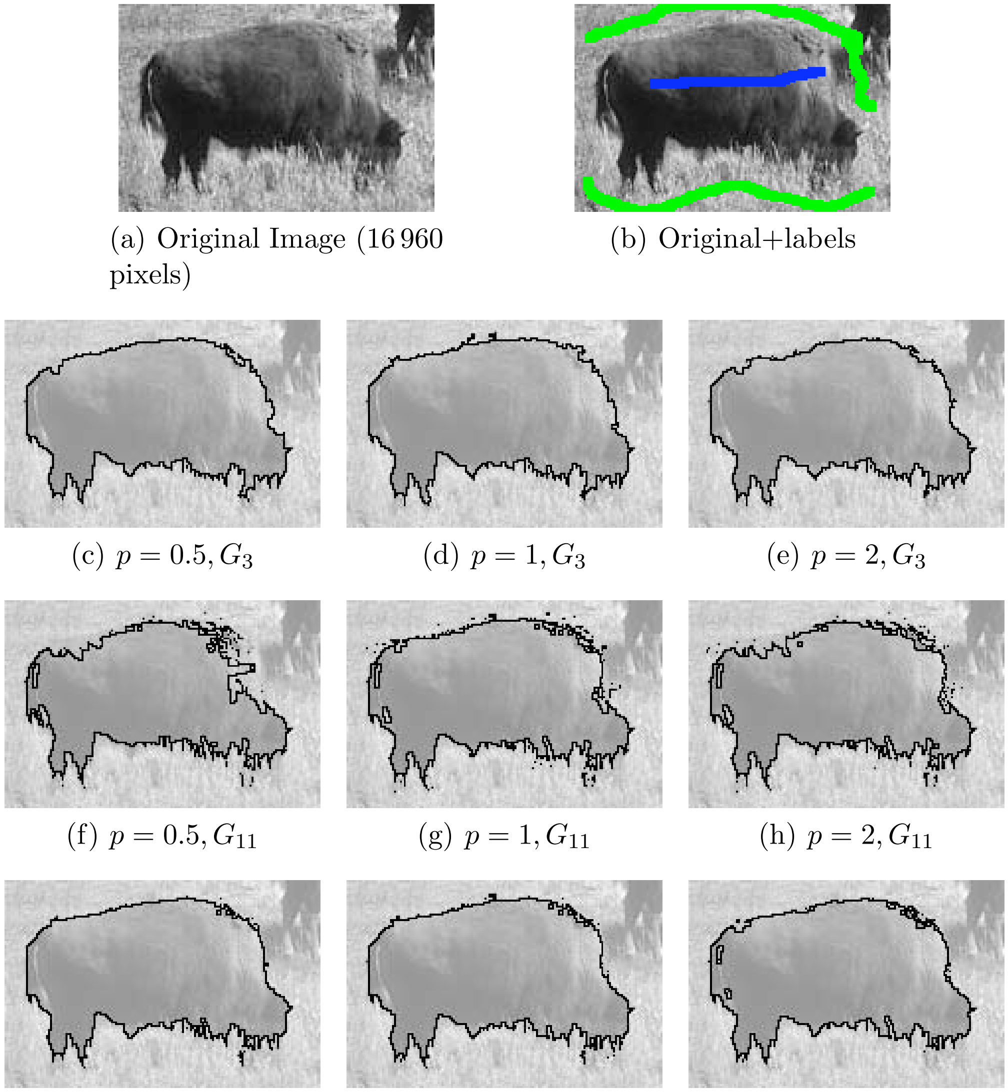
Image segmentation lies in searching relevant image regions or objects. Many successful automatic image segmentation approaches have been proposed in the literature. But, sometimes, their segmentation results are not accurate when image are more complex. Recent interactive image segmentation approaches have became increasingly popular in the image processing community. They reformulate image segmentation tasks into semi-supervised classification approaches. Thus, the segmentation is solved by label propagation strategies. We propose to use our anisotropic regularization framework to address the image semi-supervised task. Recently, similar approaches have also been proposed [2].

Image semi-supervised segmentation consists in finding objects in images by completing the user initial labels. Commonly, the user marks the desired objects to be segmented, or/and the image background. In the sequel, we directly consider the raw images without any pre-processing.

Grid graph based image semi-supervised segmentation. Fig. 4 shows the behavior of our anisotropic semi-supervised image segmentation based on grid graphs. These experiments show segmentation results for different grid graphs. Figs. 4(c), 4(d), and 4(e) is a local semi-supervised segmentation using a 8-adjacency graph (G_1). Figs. 4(f), 4(g), and 4(h) correspond to a nonlocal processing with the graph G_{11}, and the weight function (28), where $F_f(v) = f(v)$. Figs. 4(i), 4(j), and 4(k) is also a nonlocal processing with the graph G_{11}, and the weight function (28), where $F_f(v)$ is defined as a patch of size 5×5.

One can note that in the case of 8-adjacency graphs, we obtain a correct segmentation of the object. When we use a larger searching window, the object boundary are less smooth than in the local case, especially for the case where

$p < 1$. This effect is corrected by using patches as observed in the last row of Fig. 4.



(a) Original Image (16 960 pixels) (b) Original+labels

(c) $p = 0.5, G_3$ (d) $p = 1, G_3$ (e) $p = 2, G_3$

(f) $p = 0.5, G_{11}$ (g) $p = 1, G_{11}$ (h) $p = 2, G_{11}$

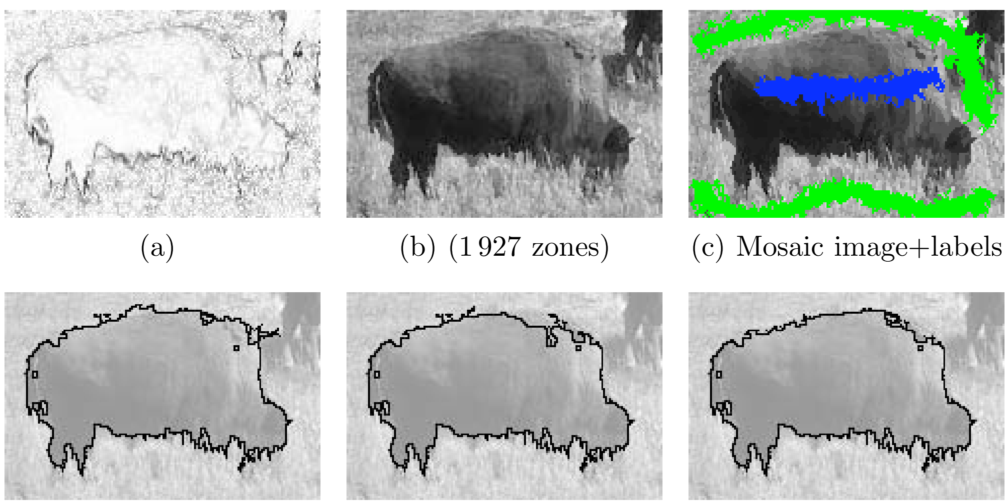
(i) $p = 0.5, G_{11}$, patch 3×3 (j) $p = 1, G_{11}$, patch 3×3 (k) $p = 2, G_{11}$, patch 3×3

Fig. 4. Grid graph based image semi-supervised segmentation. All the images were whitened in order to accentuate the user labels and the segmented boundary. (a) original image (16 960 pixels). (b) original image with user labels.

Fast image semi-supervised segmentation with simplified images. As in Fig. 4, common image semi-supervised segmentation are usually based on a pixel label propagation scheme. Their application becomes difficult when the image is of large size. To avoid this computing restriction, we can consider that image pixels are not the only relevant elements. Then, more abstract structures can be used for the segmentation process, such as image regions or "superpixels" [28]. We suggest to work, not directly with the image pixel graph representation but, with a reduced version of the image. This simplification can be thought as a graph simplification or a data reduction. To achieve this image pre-processing, any well known image pre-segmentation can be performed such as watershed techniques. In this work, this simplification is performed by an approach based on generalized Voronoï diagram (for more details see [4, 10]).

Fig. 5(b) shows the image simplification of the image of Fig. 4(a). One can note that the approach respect the main image components, notably the edge information (see the energy image Fig. 5(a)). Meanwhile, the data are significantly reduced. Fig. 5 shows the semi-supervised segmentation accomplished with a simplified version of the image Fig. 4(a). Fig. 5(b) is the simplified mosaic image with a mean color value for each zone from the original one. Fig. 5(c) is the mosaic image with the initial labeled zones. Figs. 5(d), 5(e) and 5(f) are the obtained final segmentation with several values of p. In this experiment, the initial conditions are the same as in Fig. 4, for the parameter λ and the weight function. We use the Region Adjacency Graph as the graph representation of the mosaic image, where each vertex corresponds to an image zone.

The benefits of using a simplified version of the image provides a fast segmentation scheme due to the reduced number of data to be analyze. Hence a minimal number of iterations is needed to obtain a correct segmentation.



(a) (b) (1 927 zones) (c) Mosaic image+labels

(d) $p = 0.5$ (e) $p = 1$ (f) $p = 2$

Fig. 5. Semi-supervised image segmentation based on the region adjacency graph of the mosaic image. The images are whitened for accentuate the labels and the segmented boundary. (a) energy image (b) mosaic image (1 927 zones) from Fig. 4(a) (which contains 16 690 pixels) where each zone corresponds to the mean color model of the original image. (c) mosaic image with the user label (d), (e), and (f) segmentation results obtain with the corresponding p parameter.

Nonlocal semi-supervised segmentation with simplified images. In order to illustrate the natural nonlocal expression of our anisotropic regulation framework, we perform the semi-supervised image segmentation on a fully connected graph constructed from a simplified images. Fig. 6 shows examples of this processing on color images. The graph is constructed from pre-segmented images (mosaic images of Figs. 6(b), 6(e), and 6(h)). Each vertex v corresponds to an influence zone which is described by its mean RGB color. This semi-supervised segmentation, provides a fast, simple, efficient and uncommon fully nonlocal image scheme as shown in Figs. 6(c), 6(f), and 6(i). We can quote the following interesting properties:

(1) The fully connected graph contains all the image data information on the weighted edges.
(2) Only few labels, as shown in Figs. 6(a), 6(d) and 6(g), are needed to obtain a correct segmentation result. For instance, in Fig. 6(g), the user only marks one nucleus and our fully nonlocal segmentation has found all the others (see Fig. 6(i)).
(3) The fully connected graph extends the notion of proximity between two vertices. They can be similar even if they are not spatially close or adjacent. The diffusion process allows to quickly labeled the objects in the same class even if they are not spatially adjacent, as shown in Figs. 6(g) and 6(i).
(4) The regularization process only needs a minimal number of iterations to compute a correct segmentation result. This approach provides a fast segmentation strategy.

Semi-supervised discrete data classification. To illustrate the flexibility of our anisotropic framework on non-organized data classification. Fig. 7 shows a classical example problem: the two moons. In this example the fully connected graph is computed in order to connect all the data points to each other. The initial data is composed of 99 unlabeled points in each class. Only one label per class is placed as show in Fig. 7(a). Fig. 7(b) shows the obtained result and the correctness of the classification.

7 Conclusion

In this paper, we have proposed a discrete anisotropic regularization framework on arbitrary graphs, to filter and classify discrete data. The regularization is based on the $\mathcal{L}_p$-norm of the graph gradient. It leads to semi-discrete and discrete anisotropic diffusion processes. When the data are naturally organized, such as images, we show that local and nonlocal regularization functionals have the same expression. We apply this framework to perform image and data filtering and clustering. The main ongoing work is to use the proposed anisotropic regularization functional in other discrete variational problems, such as one involves in binary segmentation.

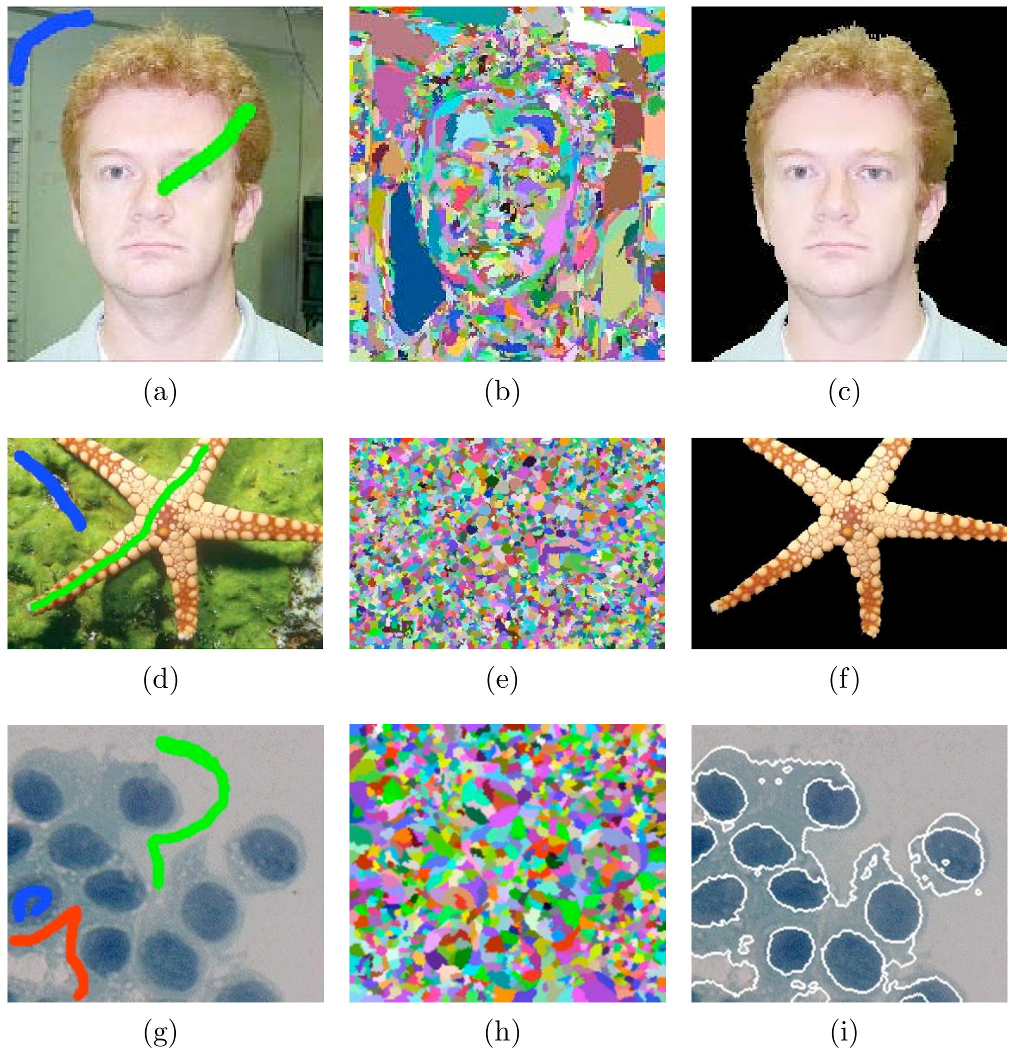


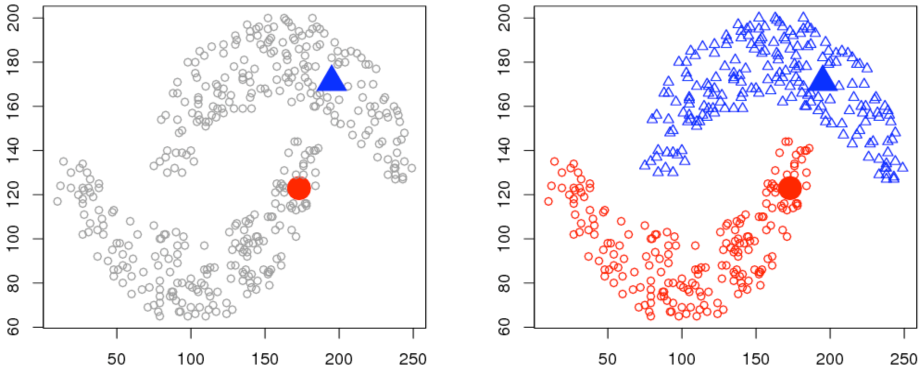
Fig. 6. Semi-supervised image segmentation. First column: original images with user initial labels. (a) and (d) Two classes problem (foreground and background). (g) Three classes problem (nuclei, cytoplasm, and background). Second column: influence zones images. Third column: segmentation results.

References

[1] L. Alvarez, F. Guichard, P.-L. Lions, J.-M. Morel, Axioms and fundamental equations of image processing, Archive for Rational Mechanics and Analysis 123 (3) (1993) 199–257.

[2] A. K. S. and Leo Grady, A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm, in: ICCV 2007, 2007.

[3] F. Andreu, J. M. Mazon, J. D. Rossi, J. Toledo, A nonlocal p-laplacian evolution equation with neumann boundary conditions, submitted, (2007).



(a) (b)

Fig. 7. Semi-supervised classification toy example (the two moons).

[4] P. A. Arbeláez, L. D. Cohen, Energy partitions and image segmentation, Journal of Mathematical Imaging and Vision 20 (1–2) (2004) 43–57.

[5] G. Aubert, P. Kornprobst, Mathematical Problems in Image Processing. Partial Differential Equations and the Calculus of Variations, vol. 147 of Applied Mathematical Sciences, 2nd ed., Springer, 2006.

[6] P. Bates, A. Chmaj, A discrete convolution model for phase transitions, Arch. Rat. Mech. Anal. 150 (1999) 281–305.

[7] M. Belkin, P. Niyogi, V. Sindhwani, Manifold regularization: A geometric framework for learning from labeled and unlabeled examples, Journal of Machine Learning Research 7 (2006) 2399–2434.

[8] A. Bensoussan, J.-L. Menaldi, Difference equations on weighted graphs, Journal of Convex Analysis 12 (1) (2004) 13–44.

[9] M. Bogoya, R. Ferreira, J. D. Rossi, Neumann boundary conditions for a nonlocal nonlinear diffusion operator. continuous and discrete models, in: Proceedings of the American Mathematical Society, vol. 135 of 12, 2007.

[10] S. Bougleux, A. Elmoataz, Image smoothing and segmentation by graph regularization, in: Bebis, Boyle, Koracin, Parvin (eds.), Advances in Visual Computing, vol. 3804 of Lecture Notes in Computer Science, Springer, 2005.

[11] S. Bougleux, A. Elmoataz, M. Melkemi, Discrete regularization on weighted graphs for image and mesh filtering, in: F. Sgallari, A. Murli, N. Paragios (eds.), Scale Space and Variational Methods in Computer Vision, vol. 4485 of LNCS, Springer, 2007.

[12] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (11) (2001) 1222–1239.

[13] A. Buades, B. Coll, J. Morel, C. Sbert, Non local demosaicing, Preprint CMLA 2007-15, submitted to IEEE TIP, (2007).

[14] A. Buades, B. Coll, J.-M. Morel, A review of image denoising algorithms, with a new one, Multiscale Modeling and Simulation 4 (2) (2005) 490–530.

[15] T. Chan, S. Osher, J. Shen, The digital TV filter and nonlinear denoising, IEEE Transactions on Image Processing 10 (2) (2001) 231–241.

[16] T. Chan, J. Shen, Image Processing and Analysis: Variational, PDE, Wavelet, and Stochastic Methods, SIAM, 2005.

[17] F. Chung, Spectral Graph Theory, No. 92 in CBMS Regional Conference Series in Mathematics, American Mathematical Society, 1997.

[18] J. Darbon, M. Sigelle, Exact optimization of discrete constrained total variation minimization problems, in: R. Klette, J. Zunic (eds.), Proceedings of the 10th Int. Workshop on Combinatorial Image Analysis, vol. 3322 of LNCS, 2004.

[19] A. Elmoataz, O. Lézoray, S. Bougleux, Nonlocal discrete regularization an weighted graphs: a framework for image and manifolds processing, IEEE Transactions on Image Processing (2007) to appear.

[20] G. Gilboa, S. Osher, Nonlocal linear image regularization and supervised segmentation, SIAM Multiscale Modeling and Simulation 6 (2) (2007) 595–630.

[21] G. Gilboa, S. Osher, Nonlocal operators with applications to image processing, Report 07-23, UCLA, Los Angeles (July 2007).

[22] C. Kervrann, J. Boulanger, P. Coupé, Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal, in: F. Sgallari, A. Murli, N. Paragios (eds.), Proc. of the 1st Int. Conf. on Scale Space and Variational Methods in Computer Vision (SSVM), vol. 4485 of LNCS, Springer, 2007.

[23] S. Kindermann, S. Osher, P. W. Jones, Deblurring and denoising of images by nonlocal functionals, SIAM Multiscale Modeling and Simulation 4 (4) (2005) 1091–1115.

[24] P. McDonald, R. Meyers, Diffusions on graphs, poisson problems and spectral geometry, Transactions of the American Mathematical Society 354 (12) (2002) 5111–5136.

[25] S. Osher, J. Shen, Digitized PDE for data restoration., in: G. Anastassiou (ed.), Handbook of Analytic Computational Methods in Applied Mathematics, 1st ed., chap. 16, Chapman&Hall/CRC, 2000, pp. 751–771.

[26] P. Perona, J. Malik, scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (7) (1990) 629–639.

[27] G. Peyré, Image processing with non-local spectral bases, Tech. rep., CEREMADE, Université Paris Dauphine. (2007).

[28] X. Ren, J. Malik, Learning a classification model for segmentation, in: Proceedings Ninth IEEE International Conference on Computer Vision, vol. 1, 2003.

[29] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Phys. D 60 (1–4) (1992) 259–268.

[30] S. Smith, J. Brady, SUSAN - a new approach to low level image processing, International Journal of Computer Vision (IJCV) 23 (1997) 45–78.

[31] C. Tomasi, R. Manduchi, Bilateral filtering for gray and color images, in: Sixth International Conference on Computer Vision (ICCV'98), IEEE Computer Society, 1998.

[32] Y. Tsai, S. Osher, Total variation and level set methods in image science, Acta Numerica 14 (2005) 509–573.

[33] D. Tschumperlé, Fast anisotropic smoothing of multi-valued images using curvature-preserving PDEś., International Journal of Computer Vision 68 (1) (2006) 65–82.

[34] T. Viéville, An unbiased implementation of regularization mechanisms, Image and Vision Computing 23 (11) (2005) 981–998.

[35] D. Zhou, B. Schölkopf, A regularization framework for learning from graph data, in: Proceedings of the ICML-2004 Workshop on Statistical Relational Learning and its Connections to Other Fields (SRL-2004), 2004.

[36] D. Zhou, B. Schölkopf, Regularization on discrete spaces, in: Patter Recognition-Proceedings of the 27th DAGM Symposium, Springer, Berlin, 2005.

[37] D. Zhou, B. Schölkopf, Discrete Regularization, Semi-Supervised Learning, section 3.13, Adaptive computation and machine learning, MIT Press, Cambridge, MA, USA, 2006, pp. 221–232.