Spatial and Anatomical Regularization of SVM: A General Framework for Neuroimaging Data

Rémi Cuingnet, Joan Alexis Glaunès, Marie Chupin, Habib Benali, Olivier Colliot, and The Alzheimer's Disease Neuroimaging Initiative

Abstract—This paper presents a framework to introduce spatial and anatomical priors in SVM for brain image analysis based on regularization operators. A notion of proximity based on prior anatomical knowledge between the image points is defined by a graph (e.g., brain connectivity graph) or a metric (e.g., Fisher metric on statistical manifolds). A regularization operator is then defined from the graph Laplacian, in the discrete case, or from the Laplace-Beltrami operator, in the continuous case. The regularization operator is then introduced into the SVM, which exponentially penalizes high-frequency components with respect to the graph or to the metric and thus constrains the classification function to be smooth with respect to the prior. It yields a new SVM optimization problem whose kernel is a heat kernel on graphs or on manifolds. We then present different types of priors and provide efficient computations of the Gram matrix. The proposed framework is finally applied to the classification of brain Magnetic Resonance (MR) images (based on Gray Matter (GM) concentration maps and cortical thickness measures) from 137 patients with Alzheimer's Disease (AD) and 162 elderly controls. The results demonstrate that the proposed classifier generates less-noisy and consequently more interpretable feature maps with high classification performances.

Index Terms—SVM, regularization, Laplacian, Alzheimer's disease, neuroimaging

1 Introduction

Recently, there has been growing interest in Support Vector Machine (SVM) methods [1], [2], [3], [4] for brain image analysis. Theses approaches allow one to capture complex multivariate relationships in the data and have been successfully applied to the individual classification of a variety of neurological and psychiatric conditions, such as Alzheimer's Disease (AD) [5], [6], [7], [8], [9], [10], fronto-temporal dementia [6], [11], autism [12], schizophrenia [13], and Parkinsonian syndromes [14]. In these approaches, with the exception of [15], the specificity of neuroimaging data is taken into account in the feature extraction but not in the classification method per se. Brain images are indeed a prototypical case of structured data whose structure is governed by the underlying anatomical and functional organization.

A lot of research has been carried out to take the structure of the data into account in SVM approaches. For example, graph and sequence kernels [16] have been proposed to classify corresponding structured data such as chemical compounds or DNA sequences [17]. On the other hand, efforts have also been made to introduce structured priors into classification of vectorial data. In the literature, three main ways have been considered in order to include priors in SVM. The first way to include prior is to directly design the kernel function [2]. Another way is to constrain the classifier function to be locally invariant to some transformations. This can be done 1) by directly engineering a kernel which leads to locally invariant SVM [18], 2) by generating artificially transformed examples from the training set to create virtual support vectors (virtual SV) [19], 3) by using a combination of both these approaches called kernel jittering [20]. The last way is to consider SVM from the regularization viewpoint [2], [21], [22], [23].

In the case of brain imaging data, defining a proper similarity measure between individuals is challenging, and the use of an irrelevant similarity would only plunge the data into a higher dimensional space. As for the locally invariant approach, it seems restricted to relatively basic transformations, which would not be adapted to anatomical knowledge. In this paper, we therefore adopt the regularization viewpoint and show that it allows modeling various types of priors.

Graphs are a natural and flexible framework to take spatial information into consideration. Voxels of a brain

- *R. Cuingnet is with Philips Research, Medisys, 33 rue de Verdun, 92156 Suresnes Cedex, France. E-mail: remi.cuingnet@philips.com.*
- *J.A. Glaunes is with MAP5 UMR 8145, Université Paris Descartes, Sorbonne Paris Cité, Paris, France, the Equipe Cogimage—CRICM (ex CNRS UPR 640—LENA), 47 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France, the Université Pierre et Marie Curie—Paris 6, Centre de Recherche de l'Institut du Cerveau et de la Moelle Épinière, UMR_S 975, Paris, France, the Inserm U975, Paris, France, the CNRS, UMR 7225, Paris, France, and the ICM—Institut du Cerveau et de la Moëlle épinière, Paris, France. E-mail: alexis.glaunes@mi.parisdescartes.fr.*
- *M. Chupin and O. Colliot are with the Equipe Cogimage—CRICM (ex CNRS UPR 640—LENA), 47 Boulevard de l'Hôpital, 75651 Paris Cedex 13, France, the Université Pierre et Marie Curie—Paris 6, Centre de Recherche de l'Institut du Cerveau et de la Moelle Épinière, UMR_S 975, Paris, France, the Inserm U975, Paris, France, the CNRS, UMR 7225, Paris, France, and the ICM—Institut du Cerveau et de la Moëlle épinière, Paris, France. E-mail: {marie.chupin, olivier.colliot}@upmc.fr.*
- *H. Benali is with Inserm UMR_S 678, Laboratoire d'Imagerie Fonctionnelle (LIF), Faculté de Médecine Pierre et Marie Curie—Site Pitié Salpêtrière 91, Boulevard de l'Hôpital, 75634 Paris Cedex 13, France. E-mail: habib.benali@imed.jussieu.fr.*

Manuscript received 29 July 2011; revised 30 Mar. 2012; accepted 12 June 2011; published online 22 June 2011.
Recommended for acceptance by G.D. Hager.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-07-0507.
Digital Object Identifier no. 10.1109/TPAMI.2012.142.

image can be considered as nodes of a graph which models the voxels' proximity. A simple graph can be the voxel connectivity (6, 18, or 26), allowing us to model the underlying image structure [24]. More sophisticated graphs can be introduced to model the specificities of brain images. Graphs can, for example, model relational anatomy by encoding the relations between brain structures [25]. They are also widely used to model brain connectivity, be it structural or functional [26].

In this paper, we propose a framework to introduce spatial and anatomical priors into SVM using regularization operators defined from a graph [2], [21]. The graph encodes the prior by modeling the proximity between image points. We also present an analogous framework for continuous data in which the graph is replaced with a Riemannian metric on a statistical manifold. A regularization operator can then be defined from the graph Laplacian or from the Laplace-Beltrami operator. By introducing this regularization operator into the SVM, the classification function is constrained to be smooth with respect to the prior. This exponentially penalizes high-frequency components with respect to the graph or to the metric. It yields a new SVM optimization problem whose kernel is a heat kernel on graphs or on manifolds. We then introduce different types of spatial and anatomical priors, and provide efficient implementations of the Gram matrix for the different cases. Note that the framework is applicable to both 3D image data (e.g., Gray Matter (GM) maps) and surface data (e.g., cortical thickness maps). We apply the proposed approach to the classification of MR brain images from patients with Alzheimer's disease and elderly controls. The present paper extends work previously published at conferences [27], [28]. Compared to the conference papers, the present paper provides a comprehensive description of all different cases, new approaches for anatomical and combined regularization in the discrete case, experiments on simulated data, more thorough evaluation on real data, updated state of the art, and proofs of some important results.

The paper is organized as follows: In the next section, we briefly present SVM and regularization operators. We then show that the regularization operator framework provides a flexible approach to model different types of proximity (Section 3). Section 4 presents the first type of regularization, which models spatial proximity, i.e., two features are close if they are spatially close. We then present, in Section 5, a more complex type of constraints, called anatomical proximity. In the latter case, two features are considered close if they belong to the same brain network; for instance, two voxels are close if they belong to the same anatomical or functional region or if they are anatomically or functionally connected (based on fMRI networks or White Matter (WM) tracts). The two types of regularization, spatial and anatomical, are then combined in Section 6. Then, in Section 7, the proposed framework is applied to the analysis of MR images using gray matter concentration maps and cortical thickness measures from 137 patients with Alzheimer's disease and 162 elderly controls from the ADNI database (www.adni-info.org). A discussion of the methods and results is presented in Section 8.

2 Priors in SVM

In this section, we first describe the neuroimaging data that we consider in this paper. Then, after some background on SVM and on how to add prior knowledge in SVM, we describe the framework of regularization operators.

2.1 Brain Imaging Data

In this contribution, we consider any feature computed either at each voxel of a 3D brain image or at any vertex of the cortical surface. Typically, for anatomical studies the features could be tissue concentration maps, such as gray matter or white matter, the 3D case, or cortical thickness maps for the surface case. The proposed methods are also applicable to functional or diffusion weighted MRI. We further assume that 3D images or cortical surfaces were spatially normalized to a common stereotaxic space (e.g., [29], [30]) as in many group studies or classification methods [7], [8], [10], [13], [31], [32].

Let $\mathcal{V}$ be the domain of the 3D images or surfaces. v will denote an element of $\mathcal{V}$ (i.e., a voxel or a vertex). Thus, $\mathcal{X} = L^2(\mathcal{V})$, the set of square integrable functions on $\mathcal{V}$, together with the canonical dot product $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ will be the *input space*.

Let $\mathbf{x}_s \in \mathcal{X}$ be the data of a given subject s. In the case of 3D images, $\mathbf{x}_s$ can be considered in two different ways: 1) Since the images are discrete, $\mathbf{x}_s$ can be considered as an element of $\mathbb{R}^d$, where d denotes the number of voxels. 2) Nevertheless, as the brain is intrinsically a continuous object, we will also consider $\mathbf{x}_s$ as a real-valued function defined on a compact subset of $\mathbb{R}^3$ or more generally on a 3D compact Riemannian manifold.

Similarly, in the surface case $\mathbf{x}_s$ can be viewed either as an element of $\mathbb{R}^d$, where d denotes the number of vertices or as a real-valued function on a 2D compact Riemannian manifold.

We consider a group of N subjects with their corresponding data $(\mathbf{x}_s)_{s\in[1,N]} \in \mathcal{X}^N$. Each subject is associated with a group $(y_s)_{s\in[1,N]} \in \{-1, 1\}^N$ (typically his diagnosis, i.e., diseased or healthy).

2.2 Linear SVM

The linear SVM solves the following optimization problem [1], [2], [4]:

$$\begin{aligned}\left(\mathbf{w}^{\text{opt}}, b^{\text{opt}}\right) = \arg \min_{\mathbf{w}\in\mathcal{X}, b\in\mathbb{R}} \frac{1}{N}\sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[\langle \mathbf{w}, \mathbf{x}_s\rangle_{\mathcal{X}} + b]) \\ + \lambda \|\mathbf{w}\|^2_{\mathcal{X}},\end{aligned} \tag{1}$$

where $\lambda \in \mathbb{R}^+$ is the *regularization parameter* and ℓ_{hinge} is the *hinge loss function* defined as

$$\ell_{\text{hinge}} : u \in \mathbb{R} \mapsto (1-u)^+.$$

With a linear SVM, the *feature space* is the same as the *input space*. Thus, when the input features are the voxels of a 3D image, each element of $\mathbf{w}^{\text{opt}} = (w_v^{\text{opt}})_{v\in\mathcal{V}}$ also corresponds to a voxel. Similarly, for the surface-based methods the elements of $\mathbf{w}^{\text{opt}}$ correspond to vertices of the cortical surface. To be anatomically consistent, if $v^{(1)} \in \mathcal{V}$ and $v^{(2)} \in \mathcal{V}$ are close according to the topology of $\mathcal{V}$, their weights in the SVM classifier, $w^{\text{opt}}_{v^{(1)}}$ and $w^{\text{opt}}_{v^{(2)}}$, respectively, should be

similar. In other words, if $v^{(1)}$ and $v^{(2)}$ correspond to two neighboring regions, they should have a similar role in the classifier function. However, this is not guaranteed with the standard linear SVM (as, for example, in [7]) because the regularization term *is not a spatial regularization*. The aim of the present paper is to propose methods to ensure that $\mathbf{w}^{\text{opt}}$ is spatially regularized.

2.3 Regularization Operators

Our aim is to introduce spatial regularization of the classifier function. This is done through the definition of a *regularization operator* P. Following [2], [21], P is defined as a linear map from a space $\mathcal{U} \subset \mathcal{X}$ into $\mathcal{X}$. When P is bijective and symmetric, the minimization problem

$$\min_{\mathbf{u}\in\mathcal{U},b\in\mathbb{R}} \frac{1}{N}\sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[\langle \mathbf{u}, \mathbf{x}_s\rangle_{\mathcal{X}} + b]) + \lambda\|P\mathbf{u}\|_{\mathcal{X}}^2 \qquad (2)$$

is equivalent to a linear SVM on the data $(P^{-1}\mathbf{x}_s)_s$:

$$\min_{\mathbf{w}\in\mathcal{X},b\in\mathbb{R}} \frac{1}{N}\sum_{s=1}^{N} \ell_{\text{hinge}}\big(y_s\big[\langle \mathbf{w}, P^{-1}\mathbf{x}_s\rangle_{\mathcal{X}} + b\big]\big) + \lambda\|\mathbf{w}\|_{\mathcal{X}}^2. \qquad (3)$$

Similarly, it can be seen as an SVM minimization problem on the raw data with kernel K defined by $K(\mathbf{x}_1, \mathbf{x}_2) = \langle P^{-1}\mathbf{x}_1, P^{-1}\mathbf{x}_2\rangle_{\mathcal{X}}$.

One has to define the regularization operator P so as to obtain the suitable regularization for the problem.

3 Laplacian Regularization

Spatial regularization requires the notion of proximity between elements of $\mathcal{V}$. This can be done through the definition of a graph in the discrete case or a metric in the continuous case. In this section, we propose spatial regularization based on the Laplacian for both of these proximity models. This penalizes the high-frequency components with respect to the topology of $\mathcal{V}$.

3.1 Graphs

When $\mathcal{V}$ is finite, weighted graphs are a natural framework to take spatial information into consideration. Voxels of a brain image can be considered as nodes of a graph which models the voxels' proximity. This graph can be the voxel connectivity (6, 18, or 26) or a more sophisticated graph.

The Laplacian matrix L of a graph is defined as $L = D - A$, where A is the adjacency matrix and D is a diagonal matrix verifying

$$D_{i,i} = \sum_j A_{i,j}$$

[33]. Note that the graph Laplacian can be interpreted, in some cases, as a discrete representation of the Laplace-Beltrami operator (e.g., as in Section 4).

We chose the following regularization operator:

$$\begin{aligned} P_\beta\colon \mathcal{U} = \mathcal{X} &\to \mathcal{X} \\ \mathbf{u} &\mapsto e^{\frac{1}{2}\beta L}\mathbf{u}. \end{aligned} \qquad (4)$$

The parameter β controls the size of the regularization. The optimization problem then becomes

$$\min_{\mathbf{w}\in\mathcal{X},b\in\mathbb{R}} \frac{1}{N}\sum_{s=1}^{N} \ell_{\text{hinge}}(y_s[\langle \mathbf{w}, \mathbf{x}_s\rangle + b]) + \lambda\|e^{\frac{1}{2}\beta L}\mathbf{w}\|^2. \qquad (5)$$

Such a regularization term exponentially penalizes the high-frequency components and thus forces the classifier to consider as similar voxels which are highly connected according to the graph adjacency matrix. Note that the eigenvectors of the graph Laplacian correspond to functions partitioning the graph into clusters. They can be considered as a soft min-cut partition of $\mathcal{V}$ [34]. As a result, such regularization operators strongly penalize the components of $\mathbf{w}$ which vary a lot over coherent clusters in the graph.

According to the previous section, the new minimization problem (5) is equivalent to an SVM optimization problem. The new kernel K_β is given by

$$K_\beta(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T e^{-\beta L}\mathbf{x}_2. \qquad (6)$$

This is a heat or diffusion kernel on a graph. Diffusion kernels on graphs were also used by Kondor and Lafferty [35] to classify complex objects which are nodes of a graph defining the distance between them. Thus, in this approach the nodes of the graph are the objects to classify, which is different from our approach where the nodes are the features. Laplacian regularization was also used in satellite imaging [36] but, again, the nodes were the objects to classify. Our approach can also be considered as spectral regularization on the graph [37].

3.2 Compact Riemannian Manifolds

In this paper, when $\mathcal{V}$ is continuous, it can be considered as a 2D (e.g., surfaces) or a 3D (e.g., 3D euclidean or more complex) compact Riemannian manifold $(\mathcal{M}, g)$, possibly with boundaries. The metric, g, then models the notion of proximity required for the spatial regularization. Such spaces are complete Riemmanian manifolds; thus, on such spaces, the heat kernel exists [38], [39]. Therefore, the Laplacian regularization presented in the previous paragraph can be extended to compact Riemannian manifolds [39].

Unlike the discrete case, for the continuous case, the regularization operator P is not defined on the whole space $\mathcal{X}$ but on a subset $\mathcal{U} \subset \mathcal{X}$. Therefore, we first need to define the domain $\mathcal{U}$. In the following, let Δ_g denote the Laplace-Beltrami operator.[1] Let $(\mathbf{e}_n)_{n\in\mathbb{N}}$ be an orthonormal basis of $\mathcal{X}$ of eigenvectors of Δ_g (with homogeneous Dirichlet boundary conditions) [38], [40]. Let $(\mu_n)_{n\in\mathbb{N}} \in \mathbb{R}^{+\mathbb{N}}$ be the corresponding eigenvalues. We define $\mathcal{U}_\beta$ for $\beta > 0$ as

$$\mathcal{U}_\beta = \left\{ \sum_{n\in\mathbb{N}} u_n\mathbf{e}_n \,|\, (u_n)_{n\in\mathbb{N}} \in \ell^2 \text{and} \big(e^{\frac{1}{2}\beta\mu_n}u_n\big)_{n\in\mathbb{N}} \in \ell^2 \right\}, \qquad (7)$$

where ℓ^2 denotes the set of square-summable sequences.

Similarly to the graphs, we chose the following regularization operator:

$$\begin{aligned} P_\beta : \quad \mathcal{U}_\beta &\to \mathcal{X} \\ \mathbf{u} = \sum_{n\in\mathbb{N}} u_n\mathbf{e}_n &\mapsto e^{\frac{1}{2}\beta\Delta_g}\mathbf{u} = \sum_{n\in\mathbb{N}} e^{\frac{1}{2}\beta\mu_n}u_n\mathbf{e}_n. \end{aligned} \qquad (8)$$

1. Note that, in euclidean space, $\Delta_g = -\Delta$ where Δ is the Laplacian operator.

The optimization problem is also equivalent to an SVM optimization problem with kernel

$$K_\beta(\mathbf{x}_1, \mathbf{x}_2) = \left\langle \mathbf{x}_1, e^{-\beta\Delta_g}\mathbf{x}_2 \right\rangle_{\mathcal{X}}. \quad (9)$$

Laferty and Lebanon [39] proposed diffusion kernels to classify objects lying on a statistical manifold. Thus, in this approach, the points of the manifold are the objects to classify. On the contrary, in our case, the points of the manifold are the features.

In this section, we have shown that that the regularization operator framework provides a flexible approach to model different types of proximity. One has now to define the type of proximity one wants to enforce. In the following sections (Sections 4 and 5), we present different types of proximity models which correspond to different types of graphs or distances: spatial and anatomical. We then combine these two regularization types in Section 6.

4 Spatial Proximity

In this section, we consider the case of regularization based on spatial proximity, i.e., two voxels (or vertices) are close if they are spatially close.

4.1 The 3D Case

When $\mathcal{V}$ are the image voxels (discrete case), the simplest option to encode the spatial proximity is to use the image connectivity (e.g., 6-connectivity) as a regularization graph. Similarly, when $\mathcal{V}$ is a compact subset of $\mathbb{R}^3$ (continuous case), the proximity is encoded by a euclidean distance. In both cases, the spatially regularized SVM will be obtained in practice by preprocessing the data with a Gaussian smoothing kernel with standard deviation $\sigma = \sqrt{\beta}$ voxels [35] before using a standard linear SVM. Therefore, the computational complexity of the Gram matrix is

$$O(Nd\log(d)).$$

4.2 The Surface Case

The connectivity graph is not directly applicable to surfaces. Indeed, the regularization would then strongly depend on the mesh used to discretize the surface. This shortcoming can be overcome by reweighing the graph with conformal weights. In this paper, we chose a different approach by adopting the continuous viewpoint: We consider the cortical surface as a 2D Riemannian manifold and use the regularization operator defined by (8). Indeed, the Laplace-Beltrami operator is an intrinsic operator and does not depend on the chosen surface parameterization. The heat kernel has already been used for cortical smoothing, for example, in [41], [42], and [43]. We will therefore not detail this part. In [41] and [42], the Finite Difference Method (FDM) or the finite element method were used. We used the implementation described in [43] and freely available for download online.[2] It uses the parametrix expansion [44] at the first order. Similarly to the Gaussian case, the diffusion parameter β sets the amount of smoothing σ^2 with the following relation: $\sigma = \sqrt{\beta}$. The computational complexity of the Gram matrix is in

$$O(N\beta d).$$

2. http://www.stat.wisc.edu/~mchung/softwares/hk/hk.html.

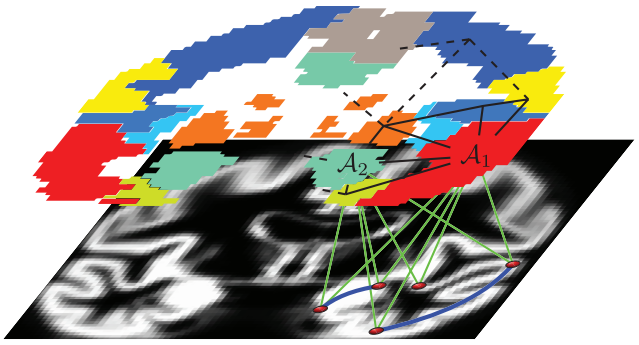

Fig. 1. Anatomical proximity encoded by a graph. The weights of the edges between nodes or voxels (red bullets) are represented by the blue arcs. They are the elements of the adjacency matrix A. They are functions of the probabilities of belonging (in green) to the regions $\mathcal{A}_r$ of an atlas (matrix E) and of the links (in black) between regions (matrix C).

5 Anatomical Proximity

In this section, we consider a different type of proximity, which we call anatomical proximity. Two voxels are considered close if they belong to the same brain network. For example, two voxels can be close if they belong to the same anatomical or functional region (defined for example by a probabilistic atlas). This can be seen as a "short-range" connectivity. Another example is that of "long-range" proximity, which models the fact that distant voxels can be anatomically (through white matter tracts) or functionally connected (based on fMRI networks).

We focus on the discrete case. The presented framework can be used either for 3D images or surfaces and computed very efficiently.

5.1 The Graph: Atlas and Connectivity

Let $(\mathcal{A}_1, \ldots, \mathcal{A}_R)$ be the R Regions of Interest (ROI) of an atlas and $p(v \in A_r)$ the probability that voxel v belongs to region $\mathcal{A}_r$. Then the probability that two voxels, $v^{(1)}$ and $v^{(2)}$, belong to the same region is: $\sum_{r=1}^{R} p((v^{(1)}, v^{(2)}) \in \mathcal{A}_r^2)$. We assume that if $v^{(1)} \neq v^{(2)}$, then

$$p\left((v^{(1)}, v^{(2)}) \in \mathcal{A}_r^2\right) = p\left(v^{(1)} \in \mathcal{A}_r\right) p\left(v^{(2)} \in \mathcal{A}_r\right).$$

Let $E \in \mathbb{R}^{d \times R}$ be the right stochastic matrix defined by

$$E_{i,r} = p\left(v^{(i)} \in \mathcal{A}_r\right). \quad (10)$$

Then, for $v^{(i)} \neq v^{(j)}$, the (i,j)th entry of the adjacency matrix $A = EE^T$ is the probability that the voxels $v^{(i)}$ and $v^{(j)}$ belong to the same regions.

For "long-range" connections (structural or functional), one can consider a positive semidefinite R-by-R matrix C with the (r_1, r_2)th entry being the probability that $\mathcal{A}_{r_1}$ and $\mathcal{A}_{r_2}$ are connected. Then, the probability that the voxels $v^{(i)}$ and $v^{(j)}$ are connected is $A_{i,j} = \sum_{r_1} E_{i,r_1} \sum_{r_2} C_{r_1,r_2} E_{j,r_2}$. Thus, the adjacency matrix becomes (Fig. 1):

$$A = ECE^T. \quad (11)$$

To compute the Gram matrix, one needs to calculate $e^{-\beta L}$, where L is the Laplacian matrix defined as $L = D - A$. The matrices D and A do not commute, which prevents us from computing the matrix exponential by multiplying $e^{-\beta D}$ and $e^{\beta A}$. We thus chose to use the normalized version of the Laplacian $\tilde{L}$:

$$\tilde{L} = I_d - D^{-\frac{1}{2}} E C E^T D^{-\frac{1}{2}}. \tag{12}$$

That is to say:

$$\tilde{L} = I_d - \tilde{E}\tilde{E}^T, \tag{13}$$

with $\tilde{E} = D^{-\frac{1}{2}} E C^{\frac{1}{2}}$. Then $e^{-\beta\tilde{L}} = e^{-\beta} e^{\beta\tilde{E}\tilde{E}^T}$.

As mentioned in Section 3, the eigenvectors of the graph Laplacian correspond to a partition of the graph into clusters. Using the unnormalized or normalized Laplacian corresponds to different types of partition: With the unnormalized Laplacian, these clusters are a min-cut segmentation of the graph, whereas with the normalized Laplacian, they are a normalized cut segmentation [34].

5.2 Computing the Gram Matrix

5.2.1 Formulation

General case. The matrix exponential can be computed by diagonalizing the normalized Laplacian. However, due to the images sizes, the direct diagonalization of the normalized graph Laplacian is computationally intractable. Nevertheless, in this case, it only comes to finding a basis of $(\ker \tilde{E}\tilde{E}^T)^{\perp}$ of eigenvectors of $\tilde{L}$. This is detailed in the following paragraph.

The matrix $\tilde{E}^T\tilde{E}$ is real symmetric. Let X be an R-by-R orthogonal matrix and Λ an R-by-R diagonal matrix such as

$$X^T \tilde{E}^T \tilde{E} X = \Lambda. \tag{14}$$

Let k be the rank of $\tilde{E}^T\tilde{E}$. Without loss of generality, we assumed that the sole nonzero components of Λ are its first k diagonal components $\Lambda_{1,1}, \ldots, \Lambda_{k,k}$. Let $\tilde{X}$ be the d-by-k matrix defined as follows: The rth column of $\tilde{X}$, $\tilde{X}_r$, is given by

$$\tilde{X}_r = \Lambda_{r,r}^{-\frac{1}{2}} \tilde{E} X_r, \tag{15}$$

where X_r denotes the rth column of X. Let $\tilde{\Lambda}$ be the k-by-k diagonal matrix defined by

$$\tilde{\Lambda}_{r,r} = 1 - \Lambda_{r,r}. \tag{16}$$

Note that $(\tilde{X}_r)_{r=1,\ldots,k}$ is an orthonormal eigenbasis of $(\ker \tilde{E}\tilde{E}^T)^{\perp}$ and that $(\Lambda_{r,r})_{r=1,\ldots,k}$ are the corresponding eigenvalues.

Then, the matrix exponential is given by

$$e^{-\beta\tilde{L}} = \tilde{X} e^{-\beta\tilde{\Lambda}} \tilde{X}^T + e^{-\beta}\big[I_d - \tilde{X}\tilde{X}^T\big]. \tag{17}$$

Special case of a binary atlas. When the atlas used to define the region is binary, in other words, when $p(v \in \mathcal{A}_r) \in \{0, 1\}$, the formulation of the matrix exponential can be more explicit than (17). Thus, the role of the regularization becomes more interpretable. Besides, it also leads to a much more efficient computation of the Gram matrix.

Let $d^{(r)}$ denote the number of voxels of region $\mathcal{A}_r$. Even if it means reindexing the voxels, we assume that the voxels are ordered by regions. In other words, we assume that the first $d^{(1)}$ voxels, $v^{(1)}, \ldots, v^{(d^{(1)})}$, belong to $\mathcal{A}_1$, then that voxels $v^{(d^{(1)}+1)}, \ldots, v^{(d^{(1)}+d^{(2)})}$ belong to $\mathcal{A}_2$, and so on. Thus, the adjacency matrix A is a block diagonal matrix verifying

$$A = \big(\mathbf{1}_{d^{(1)}}\mathbf{1}_{d^{(1)}}^T\big) \oplus \big(\mathbf{1}_{d^{(2)}}\mathbf{1}_{d^{(2)}}^T\big) \oplus \cdots \oplus \big(\mathbf{1}_{d^{(R)}}\mathbf{1}_{d^{(R)}}^T\big), \tag{18}$$

where $\mathbf{1}_{d^{(r)}}$ denotes the $d^{(r)}$-element column vector of all ones. This leads to the following matrix exponential:

$$e^{-\beta\tilde{L}} = e^{-\beta\tilde{L}^{(1)}} \oplus e^{-\beta\tilde{L}^{(2)}} \oplus \cdots \oplus e^{-\beta\tilde{L}^{(R)}}, \tag{19}$$

with, for all $r \in [1, R]$:

$$e^{-\beta\tilde{L}^{(r)}} = e^{-\beta} I_{d^{(r)}} + (1 - e^{-\beta}) \underbrace{\left[\frac{1}{d^{(r)}}\big(\mathbf{1}_{d^{(r)}}\mathbf{1}_{d^{(r)}}^T\big)\right]}_{\text{region averaging operator}}. \tag{20}$$

Note that, in the case $\beta = 0$, this is equivalent to the standard linear SVM with no anatomical regularization. In the limit case $\beta = +\infty$, this is equivalent to replacing each voxel with the average of its atlas region, such as in [31]. The cases $\beta \in \mathbb{R}^{+*}$ are intermediate cases.

5.2.2 Computational Complexity

The computation of the Gram matrix requires only: 1) the computation of $D^{-\frac{1}{2}}$, which is done efficiently since D is a diagonal matrix, 2) the diagonalization of an R-by-R matrix, which is also efficient since $R \sim 10^2$. The number of operations needed to compute the Gram matrix is in

$$O\big(NRd + R^3\big).$$

In the special case of a binary atlas, assuming that $R < d$, the computational complexity drops to $O(Nd)$.

5.2.3 Setting the Diffusion Parameter β

The proposed regularization exponentially penalizes the high-frequency components of the graph. More specifically, each component is weighted by $e^{-\beta\mu}$, where μ is the corresponding eigenvalue. In the previously described approach, the eigenvalues μ are known. Hence, the range of the diffusion parameter β can be chosen according to the range of eigenvalues μ. The specific range that we used in our experiments is given in Section 7.5.1.

The method described in this section can be directly applied to both 3D images and cortical surfaces. Unfortunately, the efficient implementation was obtained at the cost of the spatial proximity. The next section presents a combination of both anatomical and spatial proximity.

6 Combining Spatial and Anatomical Proximities

In the previous sections, we have seen how to define regularization based on spatial proximity or on anatomical proximity. In this section, we propose to combine both those proximities: first from a discrete viewpoint and then from a continuous viewpoint in which the data lies on a Riemannian manifold.

6.1 On Graphs

6.1.1 The Optimization Problem

One of the simplest options to combine the spatial and anatomical proximities would be to add up the two regularization terms. In the following, when the notation could be confusing subscripts will be added to distinguish the spatial case ($_\mathrm{s}$) from the anatomical case ($_\mathrm{a}$). For instance, L_s will refer to the Laplacian of the graph encoding spatial proximity. Similarly, L_a will refer to the Laplacian of the graph encoding the anatomical proximity. As a result, a way of combining both regularization terms is to consider the following optimization problem:

$$\begin{aligned}\left(\mathbf{w}^{\mathrm{opt}}, b^{\mathrm{opt}}\right) = \arg\min_{\mathbf{w}\in\mathcal{X}, b\in\mathbb{R}} \frac{1}{N}\sum_{s=1}^{N} \ell_{\mathrm{hinge}}\left(y_s[\langle \mathbf{w}, \mathbf{x}_s\rangle + b]\right) \\ + \lambda\left(\|e^{\frac{\beta_\mathrm{s}}{2}L_\mathrm{s}}\mathbf{w}\|^2 + \|e^{\frac{\beta_\mathrm{a}}{2}L_\mathrm{a}}\mathbf{w}\|^2\right).\end{aligned} \tag{21}$$

Note that the regularization parameters (λ) of the spatial regularization and of the anatomical regularization could have differed. We have chosen them to be equal to avoid tuning another parameter.

The sum of definite positive matrices is a definite positive matrix. Then, according to Section 2.3, (21) is an SVM optimization problem with kernel

$$K_{\beta_\mathrm{a},\beta_\mathrm{s}}(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{x}_1^T\left(e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}}\right)^{-1}\mathbf{x}_2. \tag{22}$$

6.1.2 Computing the Gram Matrix

General case. Note that, as mentioned in the previous sections, $e^{\beta_\mathrm{a}L_\mathrm{a}}\mathbf{x}_s$ and $e^{\beta_\mathrm{s}L_\mathrm{s}}\mathbf{x}_s$ can be computed efficiently. Therefore, $(e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}})^{-1}\mathbf{x}_s$ can be obtained following a conjugate gradient technique [45].

In the following, we will estimate the number of iterations needed for the conjugate gradient. If we assume that the two Laplacian matrices are normalized graph Laplacian, the condition number κ_2 of $(e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}})$ for the spectral norm is bounded by

$$\kappa_2 \leq \frac{e^{\beta_\mathrm{a}} + e^{2\beta_\mathrm{s}}}{2}. \tag{23}$$

When the computational complexity of the spatial term is proportional to β (for instance for the surface case), if $\beta_\mathrm{s} \geq \frac{1}{2}\beta_\mathrm{a}$, using the following factorization leads to a better bound on the number of iterations:

$$e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}} = e^{\frac{\beta_\mathrm{s}}{2}L_\mathrm{s}}\left(I_d + e^{\frac{-\beta_\mathrm{s}}{2}L_\mathrm{s}}e^{\beta_\mathrm{a}L_\mathrm{a}}e^{\frac{-\beta_\mathrm{s}}{2}L_\mathrm{s}}\right)e^{\frac{\beta_\mathrm{s}}{2}L_\mathrm{s}}.$$

In this case, the bound on the condition number drops to

$$\kappa_2 \leq \frac{1 + e^{\beta_\mathrm{a}}}{1 + e^{-2\beta_\mathrm{s}}}. \tag{24}$$

As a result, according to [45], the number of iterations needed to obtained a residual error η is at most

$$\left\lceil \log\left(\frac{\eta}{2}\right)\left(\log\left(\frac{\sqrt{1+e^{\beta_\mathrm{a}}} - \sqrt{1+e^{-2\beta_\mathrm{s}}}}{\sqrt{1+e^{\beta_\mathrm{a}}} + \sqrt{1+e^{-2\beta_\mathrm{s}}}}\right)\right)^{-1}\right\rceil. \tag{25}$$

For instance, if one considers the regularization with a binary atlas, the spectrum of L_a verifies: $\mathrm{Sp}(L_\mathrm{a}) = \{1\}$. For β_a ranging from 0 to 6 and for a residual error η lower than 10^{-4}, the number of iterations will not exceed 100 iterations. In practice, with our data the number of iterations did not exceed 43.

Special case of Gaussian spatial regularization. Note that, in the 3D case, if the spatial proximity is encoded with the image connectivity (6-connectivity), then L_s is diagonalizable by a symmetric orthogonal matrix Q which is the imaginary part of a submatrix of the Discrete Fourier Transform (DFT) matrix. Therefore, multiplying a vector by Q requires only $O(d\log(d))$ operations using the Fast Sine Transform. Let S be the diagonal matrix such that: $L_\mathrm{s} = QSQ$. Then, according to (17):

$$e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}} = e^{\beta_\mathrm{s}QSQ} + \tilde{X}\left[e^{\beta_\mathrm{a}\tilde{\Lambda}} - e^{\beta_\mathrm{a}}I_R\right]\tilde{X}^T + e^{\beta_\mathrm{a}}I_d. \tag{26}$$

Using the Woodbury matrix identity [45], the inverse matrix $(e^{\beta_\mathrm{a}L_\mathrm{a}} + e^{\beta_\mathrm{s}L_\mathrm{s}})^{-1}$ is equal to

$$\mathbf{D} - \mathbf{D}\tilde{X}\left[\left(e^{\beta_\mathrm{a}\tilde{\Lambda}} - e^{\beta_\mathrm{a}}I_R\right)^{-1} + \tilde{X}^T\mathbf{D}\tilde{X}\right]^{-1}\tilde{X}^T\mathbf{D}, \tag{27}$$

with

$$\mathbf{D} = Q\left(e^{\beta_\mathrm{s}S} + e^{\beta_\mathrm{a}}I_d\right)^{-1}Q.$$

In terms of computational complexity, the most costly steps are the computation of $\tilde{\Lambda}$ and the multiplication by $\mathbf{D}$. Therefore, based on the complexity of the anatomical framework, ((5.2.2)), the computational complexity of the Gram matrix is

$$O\left((N+R)d\log_2(d) + R^3\right).$$

6.1.3 Setting the Parameters β_s and β_a

The parameters β_s and β_a can be set using the previous two sections.

6.2 On Statistical Manifolds

Another way to combine spatial and anatomical information is to consider such a combination as a modification of the local topology induced by the spatial information with respect to some given anatomical priors. Since the brain is intrinsically a continuous object, it seems more interesting to describe local behaviors from the continuous viewpoint. So, in this section we proposed a single continuous framework to naturally integrate various prior information such as tissue information, atlas information, and spatial proximity. We first show that this can be done by considering the images or surfaces as elements of a statistical manifold together with the Fisher metric. We then give some details about the computation of the Gram matrix.

6.2.1 Fisher Metric

Let $v \in \mathbb{R}^3$ be some position in the image. The images are registered in a common space. Thus the true location is known up to the registration errors. Such spatial information can be modeled by a probability density function: $x \in \mathbb{R}^3 \mapsto p_{\mathrm{loc}}(x|v)$. A simple example would be $p_{\mathrm{loc}}(\cdot|v) \sim \mathcal{N}(v, \sigma^2_{\mathrm{loc}})$. It can be seen as a confidence index about the spatial location at voxel v.

We further assume that we are given an anatomical or a functional atlas $\mathcal{A}$ composed of R regions: $\{\mathcal{A}_r\}_{r=1\cdots R}$. Therefore, in each point $v \in \mathcal{V}$, we have a probability distribution $p_{\text{atlas}}(\cdot|v) \in \mathbb{R}^{\mathcal{A}}$ which informs about the atlas region in v.

As a result, in each point $v \in \mathbb{R}^3$, we have some information about the spatial location and some anatomical information through the atlas. Such information can be modeled by a probability density function $p(\cdot|v) \in \mathbb{R}^{\mathcal{A}\times\mathbb{R}^3}$. Therefore, we consider the parametric family of probability distributions

$$\mathcal{M} = \left\{p(\cdot|v) \in \mathbb{R}^{\mathcal{A}\times\mathbb{R}^3}\right\}_{v\in\mathcal{V}}.$$

In other words, in this section, instead of considering the voxels as such, each voxel is described by a probability distribution informing us about the atlas regions to which the voxel could belong and the certainty about the spatial location. In the following, we further assume that p_{loc} and p_{atlas} are independent. Thus, p verifies

$$p((\mathcal{A}_r, \mathbf{x})|v) = p_{\text{atlas}}(\mathcal{A}_r|v)p_{\text{loc}}(\mathbf{x}|v), \forall(\mathcal{A}_r, \mathbf{x}) \in \mathcal{A}\times\mathbb{R}^3.$$

In the following, we assume that p is sufficiently smooth in $v \in \mathcal{V}$ and that the Fisher information matrix is definite at each $v \in \mathcal{V}$. Then the parametric family of probability distributions $\mathcal{M}$ can be considered as a differential manifold [46]. A natural way to encode proximity on $\mathcal{M}$ is to use the Fisher metric since the Fisher information metric is invariant under reparameterization of the manifold. $\mathcal{M}$ with the Fisher metric is a Riemmanian manifold [46]. $\mathcal{V}$ is compact; therefore, $\mathcal{M}$ is a compact Riemannian manifold. For clarity, we present this framework only for 3D images, but it could be applied to cortical surfaces with minor changes. The metric tensor g is then given for all $v \in \mathcal{V}$ by

$$g_{ij}(v) = \mathbb{E}_v\left[\frac{\partial \log p(\cdot|v)}{\partial v_i}\frac{\partial \log p(\cdot|v)}{\partial v_j}\right], 1 \le i,j \le 3. \quad (28)$$

If we further assume that $p_{\text{loc}}(\cdot|v)$ is isotropic we have

$$g_{ij}(v) = g_{ij}^{\text{atlas}}(v) + \delta_{ij}\int_{u\in\mathcal{V}} p_{\text{loc}}(u|v)\left(\frac{\partial \log p_{\text{loc}}(u|v)}{\partial v_i}\right)^2 du, \quad (29)$$

where δ_{ij} is the Kronecker delta and g^{atlas} is the metric tensor when $p(\cdot|v) = p_{\text{atlas}}(\cdot|v)$.

When $p_{\text{loc}}(\cdot|v) \sim \mathcal{N}(v, \sigma_{\text{loc}}^2 I_3)$, we have

$$g_{ij}(v) = g_{ij}^{\text{atlas}}(v) + \frac{\delta_{ij}}{\sigma_{\text{loc}}^2}. \quad (30)$$

Note that the second term,

$$\frac{\delta_{ij}}{\sigma_{\text{loc}}^2},$$

ensures that the Fisher information matrix, $g_{ij}(v)$, is definite, which is necessary for the statistical model to be geometrically regular [46].

6.2.2 Computing the Gram Matrix

Equivalence with the heat equation. Once the notion of proximity is defined, one has to compute the Gram matrix. The computation of the kernel matrix requires the computation of $e^{-\beta\Delta_g}\mathbf{x}_s$ for all the subjects of the training set. The eigendecomposition of the Laplace-Beltrami operator is intractable since the number of voxels in a brain image is about 10^6. Hence, $e^{-\beta\Delta_g}\mathbf{x}_s$ is considered as the solution at time $t = \beta$ of the heat equation with the Dirichlet homogeneous boundary conditions of unknown $\mathbf{u}$:

$$\begin{cases} \dfrac{\partial \mathbf{u}}{\partial t} + \Delta_g \mathbf{u} = 0 \\ \mathbf{u}(t=0) = \mathbf{x}_s. \end{cases} \quad (31)$$

The Laplace-Beltrami operator is given by [38]

$$\Delta_g \mathbf{u} = \frac{-1}{\sqrt{\det g}}\sum_{j=1}^{3}\frac{\partial}{\partial v_j}\left(\sum_{i=1}^{3} h_{ij}\sqrt{\det g}\frac{\partial \mathbf{u}}{\partial v_i}\right),$$

where h is the inverse tensor of g.

Solving the heat equation. In this paragraph, s is fixed. To solve (31), one can use a variational approach [47]. We used the rectangular finite elements $\{\phi^{(i)}\}$ in space and the explicit finite difference scheme for the time discretization. ζ_x and ζ_t denote the space step and the time step, respectively. Let $U(t)$ denote the coordinates of $\mathbf{u}(t)$. Let U^n denote the coordinates of $\mathbf{u}(t = n\zeta_t)$ and U^0 denote those of $\mathbf{x}_s$. This leads to

$$\begin{cases} \mathbf{M}\dfrac{dU}{dt}(t) + \mathbf{K}U(t) = 0 \\ U(t=0) = U^0, \end{cases} \quad (32)$$

with $\mathbf{K}$ the stiffness matrix and $\mathbf{M}$ the mass matrix. The stiffness matrix $\mathbf{K}$ is given by

$$\mathbf{K}_{i,j} = \int_{v\in\mathcal{V}} \left\langle \nabla_{\mathcal{M}}\phi^{(i)}(v), \nabla_{\mathcal{M}}\phi^{(j)}(v)\right\rangle_{\mathcal{M}} d\mu_{\mathcal{M}}. \quad (33)$$

The mass matrix $\mathbf{M}$ is given by

$$\mathbf{M}_{i,j} = \int_{v\in\mathcal{V}} \phi^{(i)}(v)\phi^{(j)}(v) d\mu_{\mathcal{M}}. \quad (34)$$

The trapezoidal rule was used to approximate $\mathbf{K}$ and $\mathbf{M}$; in particular: $\mathbf{M}_{i,j} \approx \delta_{ij}\det g(v^{(i)})$.

The explicit finite difference scheme is used for the time discretization, thus U^{n+1} is given by

$$\mathbf{M}U^{n+1} = (\mathbf{M} - \zeta_t\mathbf{K})U^n. \quad (35)$$

ζ_x is fixed by the MRI spatial resolution. ζ_t is then chosen so as to respect the Courant-Friedrichs-Lewy (CFL) condition, which can be written in this case as

$$\zeta_t \le 2(\max\lambda_i)^{-1},$$

where λ_i are the eigenvalues of the general eigenproblem: $\mathbf{K}U = \lambda\mathbf{M}U$. Therefore, the computational complexity is in

$$O\big(N\beta(\max_i \lambda_i)d\big).$$

To compute the optimal time step ζ_t, we estimated the largest eigenvalue with the power iteration method [45].

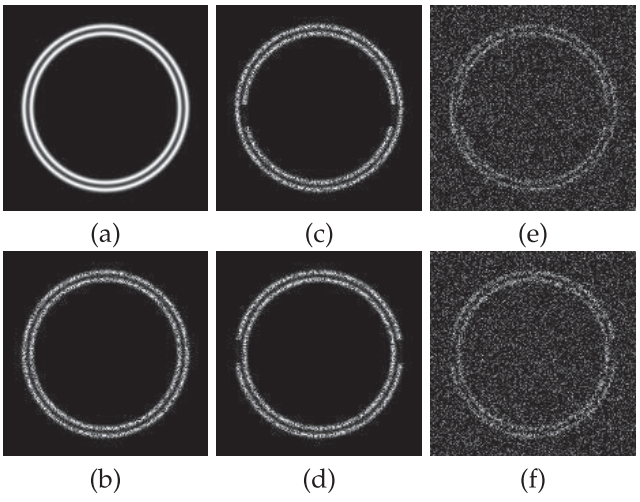

Fig. 2. Synthetic data. (a) Template composed of two concentric circles. (b) Template with multiplicative white noise. (c) Simulated lesion in the inner circle. (d) Simulated lesion in the outer circle. (e) and (f) Simulated subjects from each group after addition of white noise.

For our problem, for $\sigma_{\text{loc}} = 5$, $\lambda_{\max} \approx 15.4$ and, for $\sigma_{\text{loc}} = 10$, $\lambda_{\max} \approx 46.5$.

6.2.3 Setting the Diffusion Parameter β

We chose the same values for β as in the spatial-only case (4.1). For the metric tensor to be comparable with the spatial-only case, we normalized g with

$$\left(\frac{1}{|\mathcal{V}|}\int_{u\in\mathcal{V}}\frac{1}{3}\operatorname{tr}\left(g^{\frac{1}{2}}(u)\right)du\right)^2.$$

7 Experiments and Results

Alzheimer's disease is the most frequent neurodegenerative dementia and a growing health problem. Many group studies using structural MR images based on volumetric measurements of regions of interest (e.g., [48]), voxel-based morphometry (e.g., [48], [49]) or group comparison of cortical thickness (e.g., [50], [51]) have shown that brain atrophy in AD is spatially distributed over many brain regions. Recently, several approaches have been proposed to automatically classify patients with AD from anatomical MRI (e.g., [6], [7], [9], [10], [52], [53], [54]).

In this section, we first evaluate the proposed framework on simulated data. Then, it is applied to the analysis of MR images using gray matter concentration maps and cortical thickness measures from patients with Alzheimer's disease and elderly controls.

7.1 Simulated Data

To generate the simulated data, we constructed a template composed of two concentric circles (Fig. 2), with multiplicative white noise. From this template, we then generated two groups of 100 subjects each as follows: For each subject of the first group, we added a simulated lesion in the inner circle with angular position defined randomly between -5 and 10 degrees, and size between 1 and 4 degrees. For each subject of the second group, we added a simulated lesion in the outer circle with the same range of parameters. White noise was then added to all subjects.

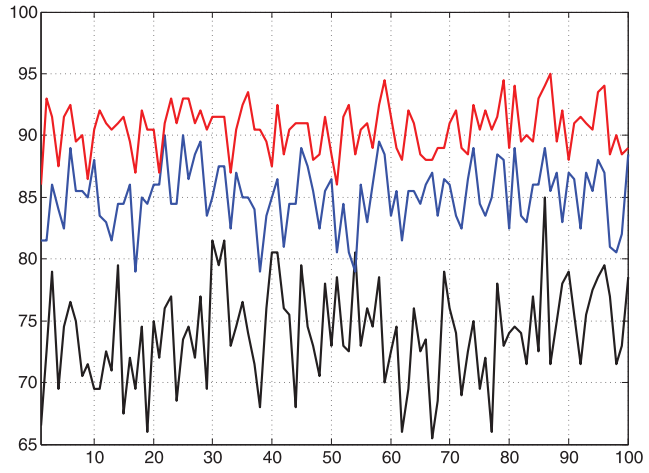

Fig. 3. Classification accuracy for each of the 100 experiments on synthetic data. In red: Combined spatial and anatomical regularization. In blue: Spatial regularization only. In black: Standard SVM.

We compared the classification performances of three different approaches: spatial regularization (with β corresponding to an FWHM of 8 mm), combined spatial and anatomical regularization (same β), no spatial regularization (standard SVM). Classification accuracy was computed using Leave-One-Out Cross-Validation (LOOCV). This experiment was repeated 100 times (i.e., we generated 100 populations of 200 subjects).

Classification performances for the three approaches are presented on Fig. 3. The combined regularization approach was consistently more accurate than the spatial regularization, which was in turn more accurate than the standard SVM.

7.2 Real Data: Material

Data used in the preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies, and nonprofit organizations as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), other biological markers, and the progression of Mild Cognitive Impairment (MCI) and early Alzheimer's disease. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the US and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research—approximately 200 cognitively normal older individuals to be followed for

TABLE 1
Demographic Characteristics of the Studied Population

Group	Diagnostic	Number	Age	Gender	MMS	# Centers
Whole set	CN	162	76.3 ± 5.4 $[60-90]$	76 M/86 F	29.2 ± 1.0 $[25-30]$	40
	AD	137	76.0 ± 7.3 $[55-91]$	67 M/70 F	23.2 ± 2.0 $[18-27]$	39
Training set	CN	81	76.1 ± 5.6 $[60-89]$	38 M/43 F	29.2 ± 1.0 $[25-30]$	35
	AD	69	75.8 ± 7.5 $[55-89]$	34 M/35 F	23.3 ± 1.9 $[18-26]$	32
Testing set	CN	81	76.5 ± 5.2 $[63-90]$	38 M/43 F	29.2 ± 0.9 $[26-30]$	35
	AD	68	76.2 ± 7.2 $[57-91]$	33 M/35 F	23.2 ± 2.1 $[20-27]$	33

Values are indicated as mean $\pm$ standard-deviation [range].

3 years, 400 people with MCI to be followed for 3 years, and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

7.2.1 Participants

We used the same study population as in [10]. We selected all the cognitively normal subjects and AD patients used in our previous paper [10]. As a result, 299 subjects were selected: 162 cognitively normal elderly controls (CN) and 137 patients with AD. Demographic characteristics of the studied population are presented in Table 1.

7.2.2 MRI Acquisition

The MR scans are T1-weighted MR images. MRI acquisition had been done according to the ADNI acquisition protocol in [55]. For each subject, we used the MRI scan from the baseline visit when available and from the screening visit otherwise. We only used images acquired at 1.5 T. To enhance standardization across sites and platforms of images acquired in the ADNI study, preprocessed images that have undergone some postacquisition correction of certain image artifacts are available [55].

7.2.3 Features Extraction

Gray matter concentration maps. For the 3D image analyses, all T1-weighted MR images were segmented into gray matter, white matter, and Cerebrospinal Fluid (CSF) using the Statistical Parametric Mapping (SPM5), London, United Kingdom, unified segmentation routine [56] and spatially normalized using the DARTEL diffeomorphic registration algorithm [30] with the default parameters. The features are the GM probability maps in the MNI space. All maps were then modulated to ensure that the overall tissue amount remains constant.

Cortical thickness. Cortical thickness measures were performed with the FreeSurfer image analysis suite (Massachusetts General Hospital, Boston, Massachusetts), which is documented and freely available for download online (http://surfer.nmr.mgh.harvard.edu/). The technical details of this procedure are described in [57] and [58].

7.3 Classification Experiments

We performed the classification of MR images from AD and controls for each regularization type presented in the previous sections. We used both the GM concentration maps and cortical thickness measures. As a result we tested the following regularization types.

7.3.1 Spatial Regularization

We tested the spatial regularization (Section 4) for both the 3D and the surface case. In the following, they will be referred to as *Voxel-Regul-Spatial* and *Thickness-Regul-Spatial*, respectively.

7.3.2 Anatomical Regularization

For the anatomical regularization (Section 5), we used the Automatic Anatomical Labeling (AAL) binary atlas [59] for the 3D case. This atlas is composed of 116 regions of interest. This approach will be referred to as *Voxel-Regul-Atlas* in the following.

As for the surface case, we used the binary cortical atlas of Desikan et al. [60]. This atlas is composed of 68 gyral-based regions of interest. This approach will be referred to as *Thickness-Regul-Atlas* in the following.

7.3.3 Combination of Spatial and Anatomical Regularization

As for the combination of the spatial and anatomical regularization described in Section 6, we tested both the graph-based (Section 6.1) and the manifold-based approaches (Section 6.2).

The graph-based approaches, *Voxel-Regul-CombineGraph* and *Thickness-Regul-CombineGraph*, also used the AAL atlas and the atlas of Desikan et al. [60] for the surface case.

We then illustrate the regularization on a statistical manifold. The atlas information used was only the tissue types. We used gray matter, white matter, and cerebrospinal fluid templates. This approach will be referred to as *Voxel-Regul-CombineFisher* in the following.

7.3.4 No Regularization

To assess the impact of the regularization we also performed two classification experiments with no regularization: *Voxel-Direct* and *Thickness-Direct*.

Namely, *Voxel-Direct* refers to an approach which consists of considering the voxels of the GM probability maps directly as features in the classification. Similarly, *Thickness-Direct* consists of considering cortical thickness values at every vertex directly as features in the classification with no other preprocessing step.

7.4 OMH Coefficient Maps

The classification function obtained with a linear SVM is the sign of the inner product of the features with $\mathbf{w}^{\mathrm{opt}}$, a vector orthogonal to the optimal margin hyperplane (OMH) [1], [2]. Therefore, if the absolute value of the ith component of

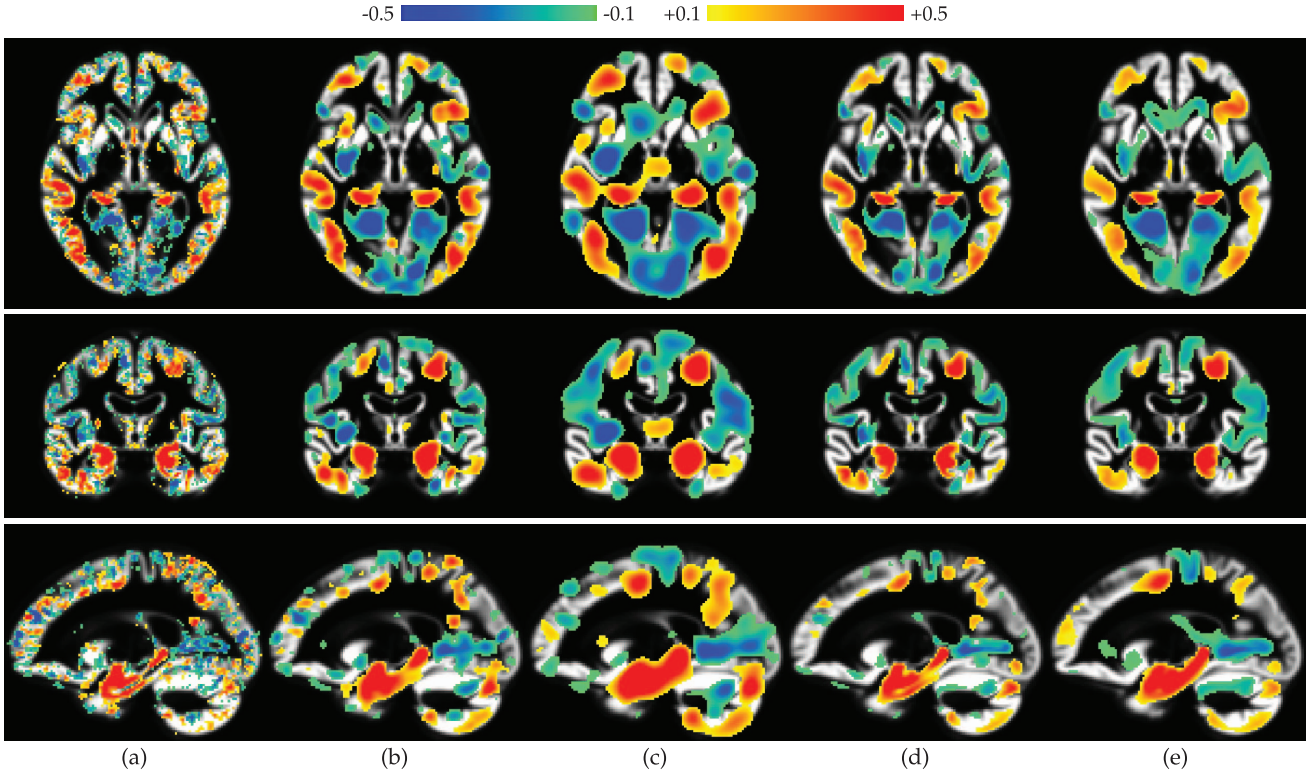

Fig. 4. Normalized $\mathbf{w}^{\text{opt}}$ coefficients for: (a) *Voxel-Direct*, (b) *Voxel-Regul-Spatial* (FWHM = 4 mm), (c) *Voxel-Regul-Spatial* (FWHM = 8 mm), (d) *Voxel-Regul-CombineFisher* (FWHM$\sim$4 mm, $\sigma_{\text{loc}} = 10$), and (e) *Voxel-Regul-CombineFisher* (FWHM$\sim$8 mm, $\sigma_{\text{loc}} = 10$).

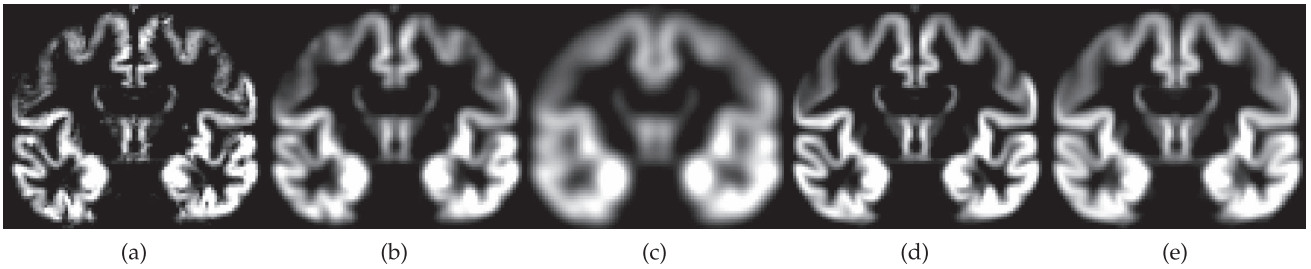

Fig. 5. Gray probability map of a control subject: (a) original map, (b) preprocessed with a 4 mm FWHM Gaussian kernel, (c) preprocessed with an 8 mm FWHM Gaussian kernel, (d)-(e) preprocessed with $e^{-\frac{\beta}{2}\Delta_g}$, where Δ_g is the Laplace-Beltrami operator of the statistical manifold and β corresponds to a 4 mm FWHM and to an 8 mm FHWM, respectively.

$\mathbf{w}^{\text{opt}}$, $|w_i^{\text{opt}}|$, is small compared to the other components $(|w_j^{\text{opt}}|)_{j\neq i}$, the ith feature will have a small influence on the classification. Conversely, if $|w_i^{\text{opt}}|$ is relatively large, the ith feature will play an important role in the classifier. Thus, the optimal weights $\mathbf{w}^{\text{opt}}$ allow us to evaluate the anatomical consistency of the classifier. In all experiments, the C parameter of the SVM was fixed to one ($\lambda = \frac{1}{2NC}$ [2]).

As an illustration of the method, we present, for some of the experiments presented in Section 7.3, the maps associated to the OMH varying the regularization parameter β. The optimal SVM weights $\mathbf{w}^{\text{opt}}$ are shown in Figs. 4 and 6. For regions in warm colors, tissue atrophy increases the likelihood of classification into AD. For regions in cool colors, it is the opposite.

7.4.1 Spatial Regularization

Fig. 4a shows the $\mathbf{w}^{\text{opt}}$ coefficients obtained with *Voxel-Direct*. When no spatial or anatomical regularization has been carried out, the $\mathbf{w}^{\text{opt}}$ maps are noisy and scattered. Figs. 4b and 4c show the results with spatial proximity for the 3D case, *Voxel-Regul-Spatial*. The $\mathbf{w}^{\text{opt}}$ map becomes smoother and spatially consistent. However, it mixes tissues and does not respect the topology of the cortex. For instance, it mixes tissues of the temporal lobe with tissues of the frontal and parietal lobes (Figs. 5b and 5c).

7.4.2 Anatomical Regularization

Fig. 6 shows the OMH coefficients obtained with *Thickness-Regul-Atlas*. When no anatomical regularization has been added ($\beta = 0$), *Thickness-Regul-Atlas* corresponds to *Thickness-Direct*. The maps are then noisy and scattered (Fig. 6a). When the amount of regularization is increased, voxels of the same region tend to be considered as similar by the classifier (Figs. 6b, 6c, and 6d). Note how the anatomical coherence of the OMH varies with β.

The regions in which atrophy increases the likelihood of being classified as AD are mainly: the hippocampus, the amygdala, the parahippocampal gyrus, the cingulum, the middle and inferior temporal gyri, and the superior and inferior frontal gyri.

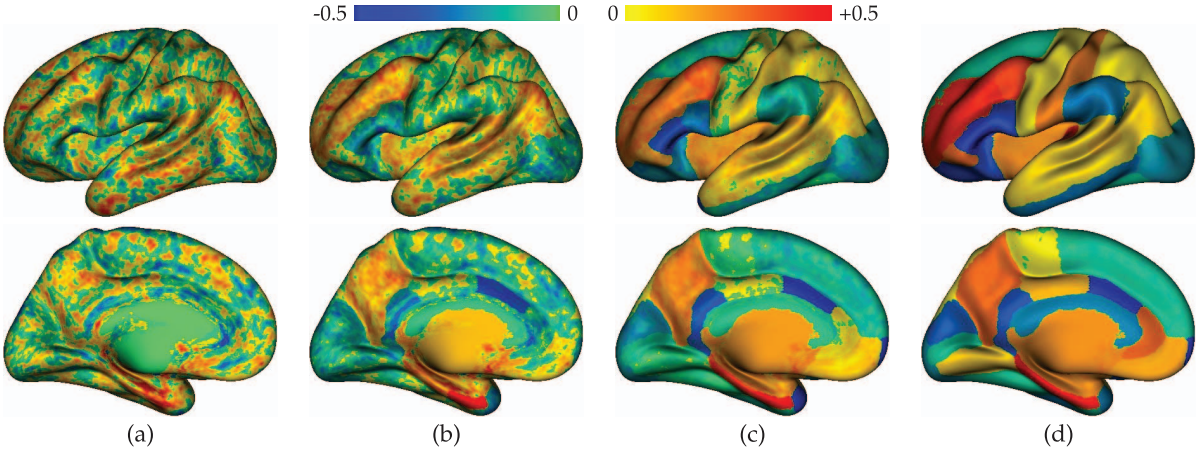

Fig. 6. Normalized $\mathbf{w}^{\text{opt}}$ coefficients for *Thickness-Regul-Atlas*. From (a) to (d), $\beta = 0, 2, 4, 6$.

7.4.3 Combining Spatial and Anatomical Regularization

The results with both spatial proximity and tissue maps, *Voxel-Regul-Fisher*, are shown in Figs. 4d and 4e. The OMH is much more consistent with the brain anatomy. Compared to *Voxel-Regul-Spatial*, it respects the topology of the cortex more (Fig. 5).

The main regions in which atrophy increases the likelihood of being classified as AD are very similar to those found with the anatomical prior: the medial temporal lobe (hippocampus, amygdala, parahippocampal gyrus), the inferior and middle temporal gyri, the posterior cingulate gyrus, and the posterior middle frontal gyrus.

We analyzed the stability of the obtained hyperplanes using bootstrap for the approach *Voxel-Regul-Fisher*. We drew 75 percent of each subject group to obtain a training set on which classification approaches were trained and the corresponding OMH estimated. The procedure was repeated 1,000 times and thus 1,000 corresponding OMH were obtained for each approach. We compared the standard SVM to the proposed regularization (with β corresponding to FWHM of 4 mm and 8 mm). For each approach, we computed the average of the 1,000 OMH. To estimate the stability of the OMH, we computed the norm of the difference between any of the 1,000 normalized OMH and the average normalized OMH. The norm was significantly lower (Student t test, $p < 0.001$) with the proposed regularization (0.64 for 8 mm, 0.65 for 4 mm) than with the standard SVM (0.68). The spatially regularized approach thus resulted in more stable hyperplanes.

7.5 Classification Performances

7.5.1 Evaluation

We assessed the classification accuracies of the different classifiers the same manner as in [10] and on the same images. As a result, in order to obtain unbiased estimates of the performances, the set of participants was randomly split into two groups of the same size: a training set and a testing set (Table 1). The division process preserved the age and sex distribution. The training set was used to determine the optimal values of the hyperparameters of each method and to train the classifier. The testing set was then only used to evaluate the classification performances. The training and testing sets were identical for all methods, except for those four cases for which the cortical thickness pipeline failed. For the cortical thickness methods, as mentioned in [10], four subjects were not successfully processed by the FreeSurfer pipeline. Those subjects could not be classified with the SVM and were therefore excluded from the training set. As for the testing set, since those subjects were neither misclassified nor correctly classified, they were considered as 50 percent misclassified.

The optimal parameter values were determined using a grid-search and leave-one-out cross validation on the training set. The grid search was performed over the ranges $C = 10^{-5}, 10^{-4.5}, \ldots, 10^{3}$ for the cost parameter of the C-SVM ($\lambda = \frac{1}{2NC}$), $\beta \in \{\alpha/\mu | \alpha \in \{0, 0.25, \ldots, 6\}, \mu \in \text{Sp}(L)\}$, $\text{FWHM} = 0, 2, \ldots, 8$ mm and $\sigma_{\text{loc}} = 5, 10$ mm.

For each approach, the optimized set of hyperparameters was then used to train the classifier using the training group; the performance of the resulting classifier was then evaluated on the testing set. In this way, we achieved unbiased estimates of the performances of each method.

7.5.2 Classification Results

The results of the classification experiments are summarized in Table 2. The accuracies ranged between 87 and 91 percent for the 3D case. The highest accuracy was obtained with *Voxel-Regul-CombineFisher* and the lowest with *Voxel-Regul-CombineGraph*. With no regularization, the classification accuracy was 89 percent.

TABLE 2
Classification Performances in Terms of Accuracies, Sensitivities (Sens), and Specificities (Spec)

Method	Accuracy	Sens	Spec
Voxel-Direct	89%	81%	95%
Voxel-Regul-Spatial	89%	85%	93%
Voxel-Regul-Atlas	90%	82%	96%
Voxel-Regul-CombineFisher	91%	88%	93%
Voxel-Regul-CombineGraph	87%	82%	90%
Thickness-Direct	83%	74%	90%
Thickness-Regul-Spatial	84%	79%	88%
Thickness-Regul-Atlas	85%	82%	86%
Thickness-Regul-CombineGraph	87%	83%	90%

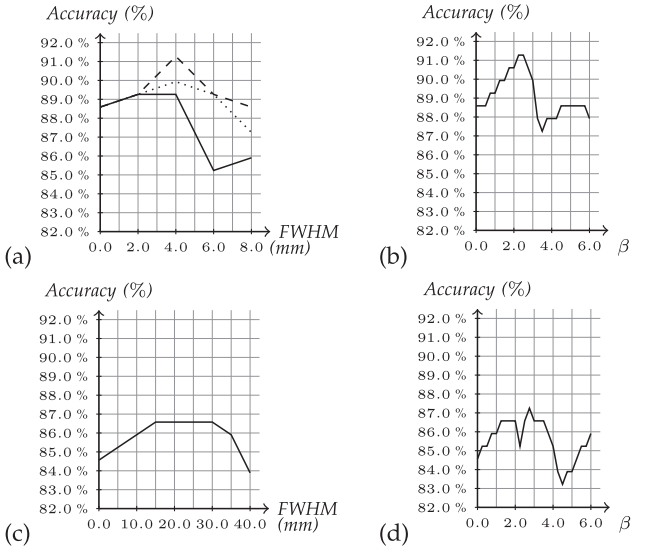

Fig. 7. Accuracies in function of the diffusion parameter ($C = 1$). (a) *Voxel-Regul-Spatial* (plain) and *Voxel-Regul-CombineFisher* (dotted: $\sigma_{\text{loc}} = 5$, dashed: $\sigma_{\text{loc}} = 10$). (b) *Voxel-Regul-Atlas*. (c) *Thickness-Regul-Spatial*. (d) *Thickness-Regul-Atlas*.

As for the surface case, *Thickness-Direct* reached 83 percent accuracy. As for the spatially and anatomically regularized approaches, the obtained accuracies were 84 and 85 percent with *Thickness-Regul-Spatial* and *Thickness-Regul-Atlas*, respectively.

On the whole there were no statistically significant differences in terms of classification accuracies, even though classification performances were slightly improved by adding spatial and/or anatomical regularization in most cases.

7.5.3 Influence of β

In this section, we further assess the influence of the β parameter on the classification performances. The C parameter of the SVM was fixed to one. The accuracies functions of β are reported in Fig. 7. It mainly yielded hump-shaped curves. The spatial or anatomical regularization improved the classification. However, a too-large amount of regularization can lead to a decrease in the classification performances.

8 Discussion

In this contribution, we proposed using regularization operators to add spatial and anatomical priors into SVM for brain image analysis. We show that this provides a flexible approach to model different types of proximity between the features. We proposed derivations for both 3D image features, such as tissue concentration maps, or surface characteristics, such as cortical thickness.

8.1 Different Proximity Models

We first considered the case of regularization based on spatial proximity, which results in spatially consistent OMH making their anatomical interpretation more meaningful. We then considered a different type of proximity model which allows modeling higher level knowledge, which we call anatomical proximity. In this model, two voxels are considered close if they belong to the same brain network. For example, two voxels can be close if they belong to the same anatomical region. This can be seen as a "short-range" connectivity. Another example is that of "long-range" proximity, which models the fact that distant voxels can be anatomically connected, through white matter tracts, or functionally connected, based on fMRI networks. This approach can be directly applied to both 3D images and cortical surfaces. Unfortunately, the efficient implementation was obtained at the cost of the spatial proximity. We thus propose to combine the anatomical and the spatial proximities. We consider two different types of formulations: a discrete viewpoint in which the proximity is encoded via a graph, and a continuous viewpoint in which the data lies on a Riemannian manifold.

8.2 Combining Spatial and Anatomical Regularization

There are two different viewpoints to combine the spatial and anatomical regularization. The first one is to consider the different types of proximity as two separate concepts. The combination can thus be done by just adding the two different regularization terms up. This is appropriate when the anatomical proximity models "long-range" proximity or connectivity. This is less appropriate for local anatomical information such as the tissue types.

Another way to combine spatial and anatomical information is to consider such a combination as a modification of the local topology induced by the spatial information with respect to some given anatomical priors. In the discrete case, local behaviors could have been encoded using histogram distances such as the χ^2 distance or the Kullback-Leibler divergence, which is in fact closely related to the Fisher metric (e.g., [39]). In this paper, we chose to describe local behaviors from the continuous viewpoint.

We propose a single framework to naturally integrate various priors such as tissue information, atlas information, and spatial proximity. In this approach, instead of considering the voxels as such, each voxel is described by a probability distribution informing us about the atlas regions to which the voxel could belong, the tissue types, and some information about the spatial location. As for the spatial information, it can be seen as a confidence index about the spatial location at each voxel and could be adapted to a specific registration algorithm. The distance between two voxels is then given by the Fisher metric.

The first limitation of this approach is that, in its current formulation, this framework is not very appropriate for binary atlases for two reasons. The first reason is the smoothness assumptions on the probability family. The second reason is the discretization. The metric tensor is evaluated at each voxel. As a result, as the norm of the metric tensor is very large at the frontier between two regions, the diffusion process is therefore stopped on a two-voxel-wide band along the frontier, which is very wide compared to the brain structures or the cortical thickness. Upsampling the image to avoid this effect is not an option due to the image size. Another limitation is that the continuous framework does not allow modeling long-range connections. This is left for future work.

8.3 Penalty Function

In this study, we forced the classifier to consider as similar voxels highly connected according to the graph adjacency matrix or close according to the given metric. This was done by penalizing the high-frequency components of the Laplacian operator. The penalty used in this study was exponential and thus led to the diffusion kernel. Many other penalty functions, such as the regularized Laplacian, the p-Step Random Walk, or the Inverse Cosine [37], for instance, could have been used instead of the diffusion process.

Nevertheless, using the diffusion process as a penalty function extends the mostly used framework which consists of smoothing the data with a Gaussian smoothing kernel as a preprocessing step.

8.4 Evaluation

Evaluation on simulated data showed that, in the presence of noise, the proposed spatial and anatomical regularization can achieve higher classification accuracies than the standard SVM. Moreover, when anatomical prior is added (using the combined regularization), one avoids mixing different regions (for example, different tissues in the context of brain imaging), which also leads to increased classification performances. Thus the proposed approach seems attractive for irregular data. On the other hand, in cases where the data would be smooth, an alternative approach could consist of placing a smoothing norm on the input space rather than on the weight space. The study of such an alternative approach is beyond the scope of this paper.

Evaluation of our approaches was then performed on 299 subjects from the ADNI database: 137 patients with Alzheimer's disease and 162 elderly controls. The results demonstrate that the proposed approach achieves high classification accuracies (between 87 and 91 percent). These performances are comparable and even often slightly higher than those reported in previously published methods for classification of AD patients. We have compared them on the same group of subjects in [10] using the same features. A linear SVM without any spatial regularization reached 89 percent accuracy. The methods *STAND-score* [8] and *COMPARE* [13] reached 81 percent accuracy and 86 percent accuracy, respectively, whereas the regularized approaches presented in this study ranged between 87 and 91 percent.

Moreover, the proposed approaches allow obtaining spatially and anatomically coherent discrimination patterns. This is particularly attractive in the context of neuroimaging in order to relate the obtained hyperplanes to the topography of brain abnormalities. In our experiments, the obtained hyperplanes were largely consistent with the neuropathology of AD, with highly discriminant features in the medial temporal lobe, as well as lateral temporal, parietal associative, and frontal areas. These areas are known to be implicated in pathological and structural abnormalities in AD (e.g., [48], [61]).

9 Conclusion

In conclusion, this paper introduces a general framework to introduce spatial and anatomical priors in SVM. Our approach allows integrating various types of anatomical constraints and can be applied to both cortical surfaces and 3D brain images. When applied to the classification of patients with Alzheimer's disease, based on structural imaging, it resulted in high classification accuracies. Moreover, the proposed regularization allows obtaining spatially coherent discriminative hyperplanes, which can thus be used to localize the topographic pattern of abnormalities associated to a pathology. Finally, it should be noted that the proposed approach is not specific to structural MRI, and can be applied to other pathologies and other types of data (e.g., functional or diffusion-weighted MRI).

Acknowledgments

This work was supported by ANR (project HM-TC, number ANR-09-EMER-006). The authors would also like to thank Olivier Druet for the useful discussions. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI; Principal Investigator: Michael Weiner; NIH grant U01 AG024904). ADNI data are disseminated by the Laboratory of Neuro Imaging at the University of California, Los Angeles. Data used in preparation of this paper were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wpcontent/uploads/how_to_apply/ADNI_Authorship_List.pdf. E-mail: remi.cuingnet@gmail.com.

References

[1] V.N. Vapnik, *The Nature of Statistical Learning Theory.* Springer-Verlag, 1995.

[2] B. Schölkopf and A.J. Smola, *Learning with Kernels.* MIT Press, 2001.

[3] J. Shawe-Taylor and N. Cristianini, *Support Vector Machines and Other Kernel-Based Learning Methods.* Cambridge Univ. Press, 2000.

[4] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis.* Cambridge Univ. Press, 2004.

[5] S. Duchesne, A. Caroli, C. Geroldi, C. Barillot, G. Frisoni, and D. Collins, "MRI-Based Automated Computer Classification of Probable AD versus Normal Controls," *IEEE Trans. Medical Imaging,* vol. 27, no. 4, pp. 509-20, Apr. 2008.

[6] C. Davatzikos, S. Resnick, X. Wu, P. Parmpi, and C. Clark, "Individual Patient Diagnosis of AD and FTD via High-Dimensional Pattern Classification of MRI," *NeuroImage,* vol. 41, no. 4, pp. 1220-1227, 2008.

[7] S. Klöppel, C.M. Stonnington, C. Chu, B. Draganski, R.I. Scahill, J.D. Rohrer, N.C. Fox, C.R. Jack, J. Ashburner, and R.S.J. Frackowiak, "Automatic Classification of MR Scans in Alzheimer's Disease," *Brain,* vol. 131, no. 3, pp. 681-689, 2008a.

[8] P. Vemuri et al., "Alzheimer's Disease Diagnosis in Individual Subjects Using Structural MR Images: Validation Studies," *NeuroImage,* vol. 39, no. 3, pp. 1186-1197, 2008.

[9] É. Gerardin et al., "Multidimensional Classification of Hippocampal Shape Features Discriminates Alzheimer's Disease and Mild Cognitive Impairment from Normal Aging," *NeuroImage,* vol. 47, no. 4, pp. 1476-1486, 2009.

[10] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic Classification of Patients with Alzheimer's Disease from Structural MRI: A Comparison of ten Methods Using the ADNI Database," *NeuroImage,* vol. 56, no. 2, pp. 766-781, 2011.

[11] S. Klöppel et al., "Accuracy of Dementia Diagnosis—A Direct Comparison between Radiologists and a Computerized Method," *Brain,* vol. 131, no. 11, pp. 2969-2974, 2008b.
[12] C. Ecker et al., "Investigating the Predictive Value of Whole-Brain Structural MR Scans in Autism: A Pattern Classification Approach," *NeuroImage,* vol. 49, no. 1, pp. 44-56, 2010.
[13] Y. Fan, D. Shen, R.C. Gur, R.E. Gur, and C. Davatzikos, "COMPARE: Classification of Morphological Patterns Using Adaptive Regional Elements," *IEEE Trans. Medical Imaging,* vol. 26, no. 1, pp. 93-105, Jan. 2007.
[14] S. Duchesne, Y. Rolland, and M. Vérin, "Automated Computer Differential Classification in Parkinsonian Syndromes via Pattern Analysis on MRI," *Academic Radiology,* vol. 16, no. 1, pp. 61-70, 2009.
[15] C. Hinrichs, V. Singh, L. Mukherjee, G. Xu, M.K. Chung, S.C. Johnson, and the ADNI, "Spatially Augmented LPboosting for AD Classification with Evaluations on the ADNI Dataset," *NeuroImage,* vol. 48, no. 1, pp. 138-149, 2009.
[16] T. Gärtner, "A Survey of Kernels for Structured Data," *ACM SIGKDD Explorations Newsletter,* vol. 5, no. 1, pp. 49-58, 2003.
[17] T. Jaakkola and D. Haussler, "Exploiting Generative Models in Discriminative Classifiers," *Proc. Conf. Advances in Neural Information Processing Systems,* pp. 487-493, 1999.
[18] B. Schölkopf, P. Simard, A. Smola, and V. Vapnik, "Prior Knowledge in Support Vector Kernels," *Proc. Conf. Advances in Neural Information Processing Systems,* pp. 640-646, 1998.
[19] B. Schölkopf, C. Burges, and V. Vapnik, "Incorporating Invariances in Support Vector Learning Machines," *Proc. Int'l Conf. Artificial Neural Networks,* pp. 47-52, 1996.
[20] D. Decoste and B. Schölkopf, "Training Invariant Support Vector Machines," *Machine Learning,* vol. 46, no. 1, pp. 161-190, 2002.
[21] A.J. Smola and B. Schölkopf, "On a Kernel-Based Method for Pattern Recognition, Regression, Approximation, and Operator Inversion," *Algorithmica,* vol. 22, no. 1/2, pp. 211-231, 1998.
[22] F. Rapaport, A. Zinovyev, M. Dutreix, E. Barillot, and J.-P. Vert, "Classification of Microarray Data Using Gene Networks," *BMC Bioinformatics,* vol. 8, no. 1, p. 35, 2007.
[23] R. Cuingnet, C. Rosso, M. Chupin, S. Lehéricy, D. Dormont, H. Benali, Y. Samson, and O. Colliot, "Spatial Regularization of SVM for the Detection of Diffusion Alterations Associated with Stroke Outcome," *Medical Image Analysis,* vol. 15, pp. 729-737, 2011.
[24] S. Geman, D. Geman, and S. Relaxation, "Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 2, pp. 721-741, Nov. 1984.
[25] I. Bloch, O. Colliot, O. Camara, and T. Géraud, "Fusion of Spatial Relationships for Guiding Recognition, Example of Brain Structure Recognition in 3D MRI," *Pattern Recognition Letters,* vol. 26, no. 4, pp. 449-57, 2005.
[26] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C. Honey, V. Wedeen, and O. Sporns, "Mapping the Structural Core of Human Cerebral Cortex," *PLoS Biology,* vol. 6, no. 7, p. e159, 2008.
[27] R. Cuingnet, M. Chupin, H. Benali, and O. Colliot, "Spatial and Anatomical Regularization of SVM for Brain Image Analysis," *Proc. Conf. Neural Information Processing Systems,* pp. 460-468, 2010.
[28] R. Cuingnet, J.A. Glaunès, M. Chupin, H. Benali, and O. Colliot, "Anatomical Regularizations on Statistical Manifolds for the Classification of Patients with Alzheimer's Disease," *Proc. Workshop Machine Learning in Medical Imaging,* pp. 201-208, 2011.
[29] D. Shen and C. Davatzikos, "HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration." *IEEE Trans. Medical Imaging,* vol. 21, no. 11, pp. 1421-1439, Nov. 2002.
[30] J. Ashburner, "A Fast Diffeomorphic Image Registration Algorithm," *NeuroImage,* vol. 38, no. 1, pp. 95-113, 2007.
[31] Z. Lao, D. Shen, Z. Xue, B. Karacali, S.M. Resnick, and C. Davatzikos, "Morphological Classification of Brains via High-Dimensional Shape Transformations and Machine Learning Methods," *NeuroImage,* vol. 21, no. 1, pp. 46-57, 2004.
[32] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Demonet, V. Duret, M. Puel, I. Berry, J.-C. Fort, P. Celsis and the ADNI, "Early Diagnosis of Alzheimer's Disease Using Cortical Thickness: Impact of Cognitive Reserve," *Brain,* vol. 132, no. 8, pp. 2036-2047, 2009.
[33] U. Von Luxburg, "A Tutorial on Spectral Clustering," *Statistics and Computing,* vol. 17, no. 4, pp. 395-416, 2007.
[34] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, Aug. 2000.
[35] R.I. Kondor and J.D. Lafferty, "Diffusion Kernels on Graphs and Other Discrete Input Spaces," *Proc. Int'l Conf. Machine Learning,* pp. 315-322, 2002.
[36] L. Gomez, G. Camps-Valls, J. Muñoz-Marí, and J. Calpe-Maravilla, "Semisupervised Image Classification with Laplacian Support Vector Machines," *IEEE Geoscience and Remote Sensing Letters,* vol. 5, no. 3, pp. 336-340, July 2008.
[37] A. Smola and R. Kondor, "Kernels and Regularization on Graphs," *Proc. 16th Ann. Conf. Computational Learning Theory,* pp. 144-158, 2003.
[38] J. Jost, *Riemannian Geometry and Geometric Analysis.* Springer, 2008.
[39] J. Lafferty and G. Lebanon, "Diffusion Kernels on Statistical Manifolds," *J. Machine Learning Research,* vol. 6, pp. 129-63, 2005.
[40] E. Hebey, *Sobolev Spaces on Riemannian Manifolds.* Springer, 1996.
[41] A. Andrade, F. Kherif, J. Mangin, K. Worsley, A. Paradis, O. Simon, S. Dehaene, D. Le Bihan, and J. Poline, "Detection of fMRI Activation Using Cortical Surface Mapping," *Human Brain Mapping,* vol. 12, no. 2, pp. 79-93, 2001.
[42] A. Cachia et al., "A Primal Sketch of the Cortex Mean Curvature: A Morphogenesis Based Approach to Study the Variability of the Folding Patterns," *IEEE Trans. Medical Imaging,* vol. 22, no. 6, pp. 754-765, June 2003.
[43] M.K. Chung, S. Robbins, and A.C. Evans, "Unified Statistical Approach to Cortical Thickness Analysis," *Information Processing in Medical Imaging,* vol. 3565, pp. 627-638, 2005.
[44] S. Rosenberg, *The Laplacian on a Riemannian Manifold: An Introduction to Analysis on Manifolds.* Cambridge Univ. Press, 1997.
[45] G. Golub and C. Van Loan, *Matrix Computations.* Johns Hopkins Univ. Press, 1996.
[46] S.-I. Amari, O.E. Barndorff-Nielsen, R.E. Kass, S.L. Lauritzen, and C.R. Rao, *Differential Geometry in Statistical Inference,* vol. 1., Inst. of Math. Statistics, 1987.
[47] O. Druet, E. Hebey, and F. Robert, *Blow-Up Theory for Elliptic PDEs in Riemannian Geometry.* Princeton Univ Press, 2004.
[48] C.D. Good, I.S. Johnsrude, J. Ashburner, R.N. Henson, K.J. Friston, and R.S. Frackowiak, "A Voxel-Based Morphometric Study of Ageing in 465 Normal Adult Human Brains," *NeuroImage,* vol. 14, no. 1, pp. 21-36, 2001.
[49] J.L. Whitwell et al., "MRI Patterns of Atrophy Associated with Progression to AD in Amnestic Mild Cognitive Impairment," *Neurology,* vol. 70, no. 7, pp. 512-520, 2008.
[50] P.M. Thompson et al., "Mapping Cortical Change in Alzheimer's Disease, Brain Development, and Schizophrenia," *NeuroImage,* vol. 23, no. Suppl 1, pp. S2-S18, 2004.
[51] J.P. Lerch, J.C. Pruessner, A.P. Zijdenbos, D.L. Collins, S.J. Teipel, H. Hampel, and A.C. Evans, "Automated Cortical Thickness Measurements from MRI Can Accurately Separate Alzheimer's Patients from Normal Elderly Controls," *Neurobiology of Aging,* vol. 29, no. 1, pp. 23-30, 2008.
[52] Y. Fan, N. Batmanghelich, C. Clark, C. Davatzikos and the ADNI, "Spatial Patterns of Brain Atrophy in MCI Patients, Identified via High-Dimensional Pattern Classification, Predict Subsequent Cognitive Decline," *NeuroImage,* vol. 39, no. 4, pp. 1731-1743, 2008.
[53] O. Colliot, G. Chételat, M. Chupin, B. Desgranges, B. Magnin, H. Benali, B. Dubois, L. Garnero, F. Eustache, and S. Lehéricy, "Discrimination between Alzheimer Disease, Mild Cognitive Impairment, and Normal Aging by Using Automated Segmentation of the Hippocampus," *Radiology,* vol. 248, no. 1, pp. 194-201, 2008.
[54] M. Chupin, E. Gerardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, and the ADNI, "Fully Automatic Hippocampus Segmentation and Classification in Alzheimer's Disease and Mild Cognitive Impairment Applied on Data from ADNI," *Hippocampus,* vol. 19, no. 6, pp. 579-587, 2009.
[55] C.R. Jack et al., "The Alzheimer's Disease Neuroimaging (ADNI): MRI Methods," *J. Magnetic Resonance Imaging,* vol. 27, no. 4, pp. 685-691, 2008.
[56] J. Ashburner and K.J. Friston, "Unified Segmentation," *NeuroImage,* vol. 26, no. 3, pp. 839-851, 2005.
[57] B. Fischl, M.I. Sereno, R.B.H. Tootell, and A.M. Dale, "High-Resolution Inter-Subject Averaging and a Coordinate System for the Cortical Surface," *Human Brain Mapping,* vol. 8, pp. 272-284, 1999.
[58] B. Fischl and A.M. Dale, "Measuring the Thickness of the Human Cerebral Cortex from Magnetic Resonance Images," *Proc. Nat'l Academy of Sciences USA,* no. 97, pp. 11044-11049, 2000.

[59] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot, "Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain," *NeuroImage,* vol. 15, no. 1, pp. 273-289, 2002.
[60] R.S. Desikan et al., "An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest," *NeuroImage,* vol. 31, no. 3, pp. 968-980, 2006.
[61] G.F. Busatto et al., "A Voxel-Based Morphometry Study of Temporal Lobe Gray Matter Reductions in Alzheimer's Disease," *Neurobiology of Aging,* vol. 24, no. 2, pp. 221-231, 2003.

Rémi Cuingnet received the BSc degree from the Ecole Polytechnique in 2006, the MSc degree in electrical engineering from Telecom ParisTech in 2007, and the PhD degree in physics from the University Paris-Sud in 2011. From 2007 to 2011, he was a research assistant within the Research Center of the Brain and Spine Institute (CR-ICM) in Paris. He is now an research scientist at Philips Research, Suresnes, France. His research interests include the area of statistical learning and medical image computing.

Joan Alexis Glaunès received the MSc degree in mathematics from ENS Cachan in 2001 and the PhD degree in mathematics from the University Paris-Nord in 2005. From 2005 to 2006, he was a postdoctoral fellow at The Johns Hopkins University and at Brown University. Since 2006, he has been an assistant professor in the Mathematics Department of the University Paris Descartes. His research interests include shape modeling, image deformations, and their applications to computational anatomy.

Marie Chupin received the BSc degree from the Ecole Supérieure d'Optique in 2000 and the PhD degree in physics from the University Paris-Sud in 2004. From 2005 to 2007, she was a postdoctoral fellow at University College London. She is now a CNRS research engineer within the Research Center of the Brain and Spine Institute (CR-ICM) in Paris. Her research interests include the area of image segmentation and medical image computing.

Habib Benali has been the director of Unit 678, Laboratory of Functional Imaging of the French National Institute of Health and Medical Research (INSERM) and Paris 6 University (UPMC) since 2008. He has been the director of the International Laboratory of Neuroimaging and Modelisation of the INSERM-UPMC and Montreal University since 2007. His current research group interests include human brain mapping and functional connectivity analysis using multimodal analysis of electromagnetic and hemodynamic processes in the brain and spinal cord.

Olivier Colliot received the MSc degree in computer science from the University of Caen in 2000, the PhD degree in signal processing from Telecom ParisTech in 2003, and the Habilitation from the University Paris-Sud in 2011. From 2003 to 2005, he was a postdoctoral fellow at McGill University. Since 2006, he has been a CNRS researcher within the Research Center of the Brain and Spine Institute (CR-ICM). His research interests include the area of shape modeling, statistical learning, and their applications to neuroimaging.