Topological Characterization of Signal in Brain Images

University of Wisconsin, Madison
Department of Biostatistics and Medical Informatics

Technical Report 207

Moo K. Chung

[1] Department of Biostatistics and Medical Informatics
[2]Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin, Madison

mkchung@wisc.edu

The technical report consists of two parts. The first part is titled *Persistence Diagrams of Cortical Surface Data* and written by Moo K. Chung, Peter Bubenik, Peter T. Kim. The second part is titled *Topological Characterization of Signal in Brain Images Using Min-max Diagrams* and written by Moo K. Chung, Vikas Singh, Peter T. Kim, Kim M. Dalton, Richard J. Davidson.

Part I. Persistence Diagrams of Cortical Surface Data

Moo K. Chung[1,2], Peter Bubenik[3], Peter T. Kim[4]

[1]Department of Biostatistics and Medical Informatics
[2]Waisman Laboratory for Brain Imaging and Behavior
University of Wisconsin, Madison, WI 53706, USA
[3]Department of Mathematics
Cleveland State University, Cleveland, Ohio 44115, USA
[4] Department of Mathematics and Statistics
University of Guelph, Guelph, Ontario N1G 2W1, Canada

mkchung@wisc.edu

Abstract. We present a novel framework for characterizing signals in images using techniques from computational algebraic topology. This technique is general enough for dealing with noisy multivariate data including geometric noise. The main tool is persistent homology which can be encoded in persistence diagrams. These are scatter plots of paired local critical values of the signal. One of these diagrams visually shows how the number of connected components of the sublevel sets of the signal changes. The use of local critical values of a function differs from the usual statistical parametric mapping framework, which mainly uses the mean signal in quantifying imaging data. Our proposed method uses all the local critical values in characterizing the signal and by doing so offers a completely new data reduction and analysis framework for quantifying the signal. As an illustration, we apply this method to a 1D simulated signal and 2D cortical thickness data.

1 Introduction

In neuroimaging, it is usually assumed that measurements f in images follow the familiar signal plus noise framework

$$f(x) = \mu(x) + \epsilon(x),\ x \in \mathbb{M} \subset \mathbb{R}^d, \tag{1}$$

where μ is the unknown mean signal, to be estimated, and ϵ is noise [5] [15] [19] [20] [26] [37]. The unknown signal is usually estimated by various spatial image smoothing techniques over $\mathbb{M}$. The most widely used smoothing technique is kernel smoothing and its variants because of their simplicity, and because they provide the theoretical context for scale spaces and Gaussian random field theory [33] [37].

In the usual statistical parametric mapping framework [15] [20] [37], inference on the model (1) proceeds as follows. If we denote an estimate of the signal by

$\widehat{\mu}$, the residual $f - \widehat{\mu}$ gives an estimate of the noise. One then constructs a test statistic $T(x)$, corresponding to a given hypothesis about the signal. As a way to account for spatial correlation of the statistic $T(x)$, the global maximum of the test statistic over the search space $\mathbb{M}$ is taken as the subsequent test statistic. Hence a great deal of the neuroimaging and statistical literature, have been devoted to determining the distribution of $\sup_{x \in \mathbb{M}} T(x)$ using random field theory [35] [37], permutation tests [31] and the Hotelling–Weyl volume of tubes calculation [30].

The use of the mean signal is one way of performing data reduction, however, this may not necessarily be the best way to characterize complex multivariate imaging data. Thus instead of using the mean signal, in this paper we propose to use what is known as persistent homology, which pairs the local critical values [12] [13] [40]. It is intuitive that the local critical values of $\widehat{\mu}$ approximately characterizes the shape of the continuous signal μ using only a finite number of scalar values. By pairing these local critical values in a nonlinear fashion and plotting them, one constructs the persistence diagram [7] [12] [29] [39].

Although persistent homology is popular in computational algebraic topology with applications in protein structure analysis [32], gene expression [11], and sensor networks [9], as far as the authors are aware, there is no such application in medical image analysis. This is the first paper that applies the concept of persistent homology to medical imaging data. The proposed method is illustrated using both simulated and real neuroimaging data. For the simulation, we use 1D Gaussian noise in (1). The 2D neuroimaging data comes from an MRI autism study where the interest is in quantifying the abnormal cortical thickness pattern in autistic subjects if there is any.

2 Persistence Diagrams

A function is called a Morse function if all critical values are unique and non-degenerate, i.e. the Hessian does not vanish [28]. We note that for integer valued digital images, critical values of intensity may not be all unique; however, the underlying continuous signal μ in (1) is likely and assumed to be a Morse function. We estimate the signal using a kernel function and obtain a smooth estimate.

For illustrative purposes, we will show how to construct the persistence diagram for a 1D Morse function. Assuming μ is a Morse function with a finite number of critical values, define a sublevel set $R(y) = \mu^{-1}(-\infty, y]$. The sublevel set is the subset of $\mathbb{R}$ that satisfies $\mu(x) \leq y$. The sublevel set can have many disjoint components. Let $\#R(y)$ be the number of connected components in the sublevel set. Let us denote the local minimums as $g_1, \cdots, g_m$ and the local maximums as $h_1, \cdots, h_n$. Since the critical values of the Morse function are all unique, we can strictly order the local minimums from the smallest to the largest as

$$g_{(1)} < g_{(2)} < \cdots < g_{(m)}$$

and similarly for the local maximums as

$$h_{(1)} < h_{(2)} < \cdots < h_{(n)}.$$

We further collect all the critical values,

$$z_1 = g_1, \ldots, z_m = g_m, z_{m+1} = h_1, \ldots, z_{m+n} = h_n$$

and order them as

$$z_{(1)} < z_{(2)} < \cdots < z_{(m+n)}.$$

At each minimum, we have the birth of a new component, i.e.

$$\#R(g_i) = \#R(g_i - \varepsilon) + 1$$

for sufficiently small ε. The new component is identified with the local minimum g_i. Similarly at each maximum, we have the death of a component, i.e.

$$\#R(h_i) = \#R(h_i - \varepsilon) - 1,$$

and two components will merge as one. The number of connected components will only change if we pass through critical points and we can iteratively compute $\#R$ at each critical value as

$$\#R(z_{(i+1)}) = \#R(z_{(i)}) \pm 1.$$

The sign depends on whether $z_{(i+1)}$ is a maximum (-1) or a minimum $(+1)$. This is the basis of Morse theory [28] that says the topological characteristics of a topological spaces are characterized by the local behavior at critical points of a Morse function on that space. Persistent homology produces pairs (g_i, h_j) of critical values so that a component is born at g_i and dies at h_j. Of course these are the (topological) parameters of interest which are unknown and to be statistically estimated with data generated according to (1).

As an example, the birth and death processes are illustrated in Figure 1, where the gray dots are simulated with Gaussian noise with mean 0 and variance 0.2^2 as

$$f(x) = \mu(x) + N(0, 0.2^2)$$

with signal $\mu(t) = 10(t - 1/2)^2 + \cos(7\pi t)/2$. The signal μ is estimated using heat kernel smoothing [5] and plotted as the red line. Now we increase y from $-\infty$ to ∞. When we hit the first critical value $y = a$, the sublevel set consists of a single point, i.e. $\widehat{\#R(a)} = 1$. When we hit the minimum at $y = b$, we have the birth of a new component at b, i.e. $\widehat{\#R(b)} = 2$. When we hit the maximum at $y = c$, the two components identified by a and c are merged together to form a single component, i.e. $\widehat{\#R(c)} = 1$.

When we pass through a maximum and merge two components, we pair the maximum with the higher of the two minimums of the two components [12]. Doing so we are pairing the birth of a component to its death. Obviously the paired extremes do not have to be adjacent to each other. If there is a boundary, the function value evaluated at the boundary is treated as a critical value. In our simulated example, we need to pair (b, c) and (d, e). Other critical values are paired similarly. The reduced persistence diagram is then the scatter plot of these pairings. For technical reasons, the persistence diagram also include all of the points (a, a), where $a \in \mathbb{R}$.

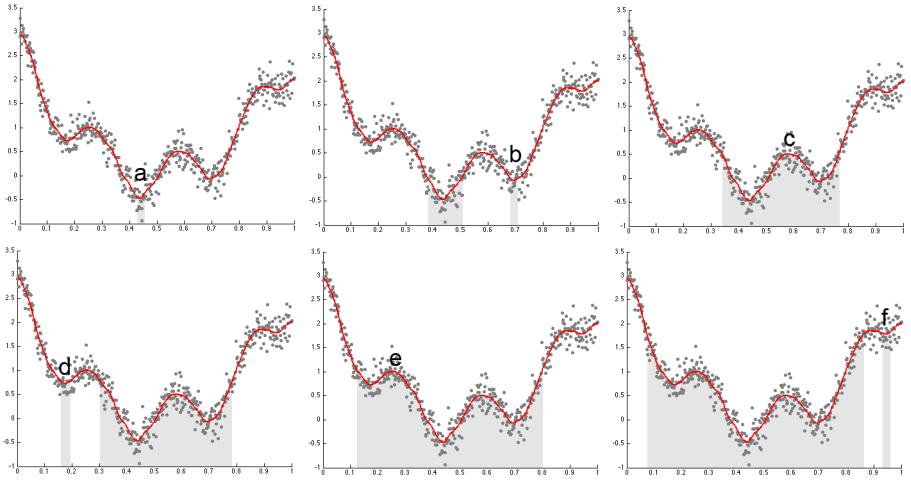

Fig. 1. The births and deaths of components in sublevel sets. We have critical values a, b, c, d, e, f, where $a < b < d < f$ are minimums and $c < e$ are maximums. At $y = a$, we have a single component marked by a single gray area. When we increase the level to $y = b$, we have the birth of a new component in addition to the existing component born at a. At the maximum $y = c$, the two components merge together to form a single component. Following the pairing rule [12], we pair (b, c) and (d, e). Other critical values are paired similarly.

2.1 Persistence Diagram for Cortical Data

For a 2D Morse function defined on a cortical manifold $\mathbb{M} \subset \mathbb{R}^3$, we need to also consider saddle points so the situation is more complicated. At a saddle point, we can have two possible pairings corresponding to either birth or death. A saddle point may join two components. This case is analogous to the local maximum in the 1D case. In this case, persistent homology pairs the value of the saddle point with the larger of the minimums of the two components. This pair is recorded in the persistence diagram of degree 0 (Figure 4). If the saddle point does not join to disconnected components, then a hole is born in the sublevel sets. Persistent homology pairs the value at this saddle point with the value of the local maximum where this hole disappears. This pair is recorded in the persistence diagram of degree 1 (Figure 4). A precise definition is given in Section 4.

Among various cortical measures, in this paper we consider cortical thickness, which has been used in characterizing various clinical populations [6] [14] [23] [24] [27] [38]. High resolution magnetic resonance images of age-matched right-handed males (6 high functioning autistic and 11 normal controls) were obtained using a 3-Tesla GE SIGNA scanner. The collected images went through intensity nonuniformity correction [34] and were spatially normalized into the MNI stereotaxic space *via* a global affine transformation [8]. Subsequently a supervised

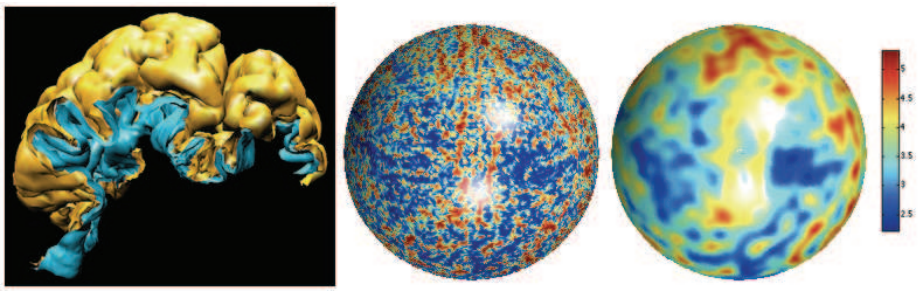

Fig. 2. Cortical thickness is computed as the distance between the outer (yellow) and the inner cortical (blue) surfaces. The cortical thickness is mapped onto a unit sphere and goes through heat kernel smoothing [5]

neural network classifier was used for tissue segmentation [22]. Brain substructures such as the brain stem were removed to make both the outer and the inner surfaces to be topologically equivalent to a sphere. A deformable surface algorithm [25] was used to obtain the inner cortical surface by deforming from a spherical mesh (Figure 2). Then the outer surface was obtained by deforming the inner surface. The deformation process establishes the structural correspondence between the two surfaces. The cortical thickness f is then defined as the distance between the corresponding vertices along the cortical mesh $\mathbb{M}$.

Since the deformable surface algorithm starts with a spherical mesh, there is no need to use other available surface flattening algorithms [1] [2] [16] [17] [36] for mapping thickness to the unit 2–sphere S^2. Let $\zeta : \mathbb{M} \to S^2$ be a sufficiently smooth surface flattening obtained from the deformable surface algorithm. Then the pullback $(\zeta^{-1})^* \widehat{\mu} = \widehat{\mu} \circ \zeta^{-1}$ projects the cortical thickness from the cortical surface $\mathbb{M}$ to the unit sphere. Figure 2 shows the pull back and the corresponding heat kernel smoothing on S^2. Note that in the process of flattening, the critical values do not change so the persistence diagram should be identical for $\widehat{\mu}$ and its pullback $(\zeta^{-1})^* \widehat{\mu}$. Therefore, we will construct the persistence diagram on the unit 2–sphere by projecting the cortical data to the sphere.

3 Kernel Smoothing

As described in Section 2.1, after the application of a deformable surface algorithm, our data is on the unit 2– sphere, S^2. So our measurement, $f : S^2 \to \mathbb{R}$ is given by the nonparametric regression formula (1), where μ is the unknown signal and ϵ is the noise. In this section, we estimate the persistent homology of the sublevel sets of $\widehat{\mu}$, an estimator of μ.

We begin by smoothing the data using the kernel,

$$K_{x_0}(x) = \max(1 - \kappa \arccos(x_0' x), 0),$$

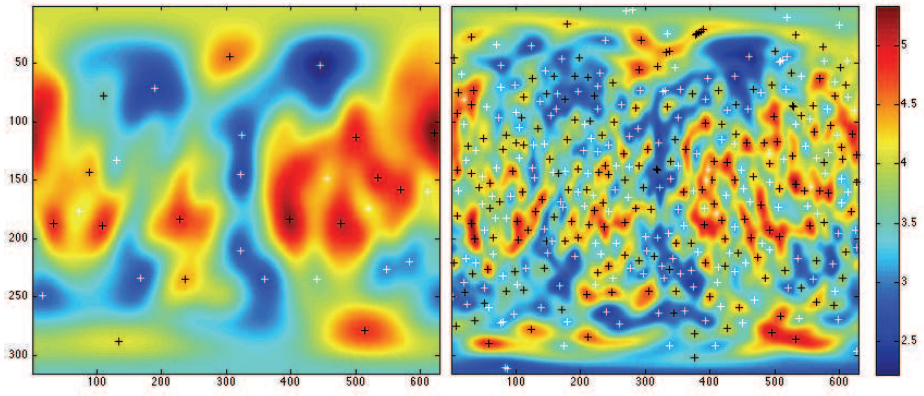

Fig. 3. The flat maps of cortical thickness at different smoothing scales. The maximums and minimums are denoted with black and white crosses respectively. The smoothing is done along the unit sphere and flattened using the Eulger angles θ (vertical axis) and φ (horizontal axis) associated with the 2-sphere. Smoother thickness produces less number of critical points and, in turn, less number of pairings.

where κ is given in [21] and $\arccos(x'y)$ gives the geodesic distance between x and y on the unit sphere. We smooth the data using the usual kernel function estimator

$$\widehat{\mu}(x) = \frac{\sum_i f(x_i) K_{x_i}(x)}{\sum_i K_{x_i}(x)}. \tag{2}$$

To implement this we need to choose the corresponding design points which we do in the following way. We start by choosing a triangulation, $\mathcal{T}$, of the sphere whose number of vertices satisfies the conditions in [3]. For our data, we start with an icosahedron and iteratively subdivide it three times, obtaining a triangulation with 1280 faces and 642 vertices.

For a sample of size n, define the estimator $\widehat{\mu}_n$ in the following way. For each vertex v in our triangulation, we define $\widehat{\mu}_n(v) = \widehat{\mu}(v)$ according to (2). For each face in our triangulation, we define $\widehat{\mu}_n$ on the face by affine interpolation from the values on the vertices. This construction is well defined on the edges, and defines a function on the sphere.

3.1 The persistence diagrams of $\widehat{\mu}_n$

It remains to calculate the persistence diagrams of the sublevel sets of $\widehat{\mu}_n$. We will see that because of the way $\widehat{\mu}_n$ is constructed, we can calculate its persistence diagrams using our triangulation, $\mathcal{T}$.

We filter $\mathcal{T}$ using $\widehat{\mu}_n$ as follows. Let $r_1 \leq r_2 \leq \ldots \leq r_m$ be the ordered list of values of $\widehat{\mu}_n$ on the vertices of the triangulation. For $1 \leq i \leq m$, let $\mathcal{T}_i$ be the subcomplex of $\mathcal{T}$ containing all vertices v with $\widehat{\mu}_n(v) \leq r_i$ and all edges whose

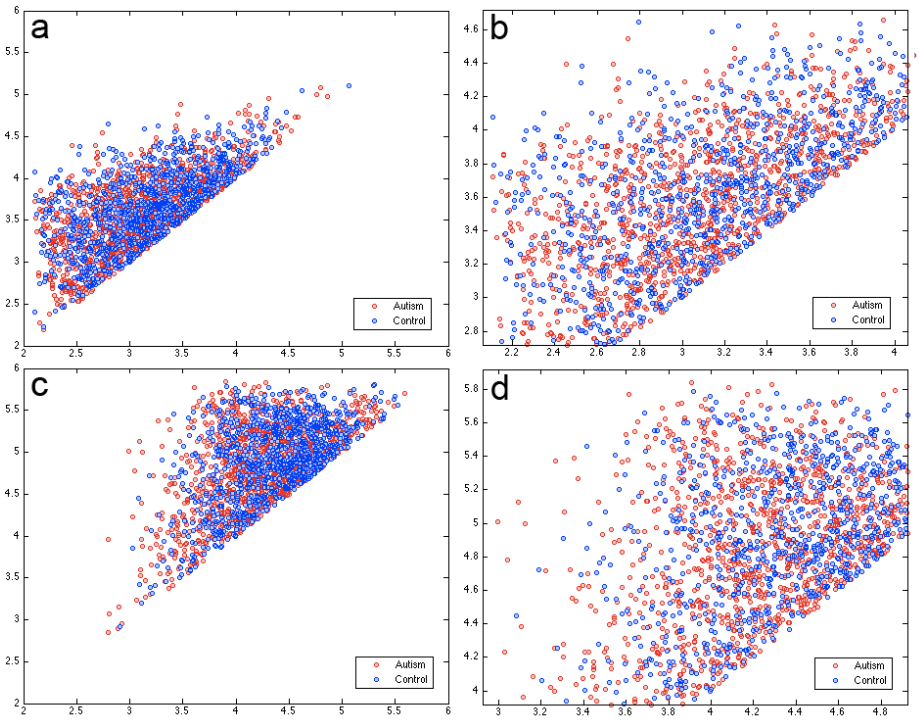

Fig. 4. The persistence diagrams for 11 control (blue) and 16 autistic (red) subjects in degree 0, (a) and (b), and degree 1, (c) and (d). One notices an additional layer of structure in the autistic group in both persistence diagrams. The figure clearly demonstrates the feasibility of using the persistence diagram for discriminating populations.

boundaries are in $\mathcal{T}_i$ and all faces whose boundaries are in $\mathcal{T}_i$. This construction is called a Vietoris–Rips complex. We obtain the following filtration of $\mathcal{T}$,

$$\phi = \mathcal{T}_0 \subset \mathcal{T}_1 \subset T_2 \subset \cdots \subset \mathcal{T}_m = \mathcal{T}.$$

The end result is that the topological properties of the sublevel sets of $\widehat{\mu}_n$ will equal the topological properties of the above filtration of $\mathcal{T}$.

Using the software Plex, [10], we calculate the persistent homology, in degrees 0, 1 and 2 of the triangulation $\mathcal{T}$ filtered according to the estimator for each of the 27 subjects. Since the data is two–dimensional, we do not expect any interesting homology in higher degrees. In degree two, the persistent homology consists of a single persistence pair (a, ∞), where a is the maximum of $\widehat{\mu}_n$.

To compare the autistic subjects and control subjects, we take the union of the persistence diagrams of the subjects (Figure 4).

4 Statistical Properties of Persistence Diagram

In this section we will make more precise definition of a persistence diagram [7] and present results that compare the topological parameters and their estimators [3] [4].

The persistent homology of the signal, μ, is encoded in its reduced persistence diagram, $\bar{D}(\mu)$, which is a multiset of points each corresponding to the persistence of one topological feature, as in the examples above. In order to define a metric for such diagrams, it is convenient to add the ordered pairs (a, a) for all $a \in \mathbb{R}$, each with infinite multiplicity. Call this multiset the *persistence diagram* of μ, denoted $D(\mu)$. We now give the precise definition.

Let k be a nonnegative integer. Given $\mu : S^2 \to \mathbb{R}$ and $a \leq b \in \mathbb{R}$ the inclusion of sublevel sets $i_a^b : S^2_{\mu \leq a} \hookrightarrow S^2_{\mu \leq b}$ induces a map on homology

$$H_k(i_a^b) : H_k(S^2_{f \leq a}) \to H_k(S^2_{f \leq b}).$$

The image of $H_k(i_a^b)$ is the persistent homology group from a to b. Let β_a^b be its dimension. This counts the independent homology classes which are born by time a and die after time b.

Call a real number a a *homological critical value* of μ if for all sufficiently small $\varepsilon > 0$ the map $H_k(i_{a-\varepsilon}^{a+\varepsilon})$ is not an isomorphism. Call μ *tame* if it has finitely many homological critical values, and for each $a \in \mathbb{R}$, $H_k(S^2_{\mu \leq a})$ is finite dimensional. In particular, any Morse function on a compact manifold is tame.

Assume that μ is tame. Choose ε smaller than the distance between any two homological critical values. For each pair of homological critical values $a < b$, we define their *multiplicity* μ_a^b which we interpret as the number of independent homology classes that are born at a and die at b. We count the homology classes born by time $a + \varepsilon$ that die after time $b - \varepsilon$. Among these subtract those born by $a - \varepsilon$ and subtract those that die after $b + \varepsilon$. This double counts those born by $a - \varepsilon$ that die after $b + \varepsilon$, so we add them back. That is,

$$\mu_a^b = \beta_{a+\varepsilon}^{b-\varepsilon} - \beta_{a-\varepsilon}^{b-\varepsilon} - \beta_{a+\varepsilon}^{b+\varepsilon} + \beta_{a-\varepsilon}^{b+\varepsilon}.$$

The *reduced persistence diagram* of μ, $\bar{D}(\mu)$, is the multiset of pairs (a, b) together with their multiplicities μ_a^b. We call this a diagram since it is convenient to plot these points on the plane. We will see that it is useful to add homology classes which are born and die at the same time. Let the *persistence diagram* of μ, $D(\mu)$, be given by the union of $\bar{D}(\mu)$ and $\{(a, a)\}_{a \in \mathbb{R}}$ where each (a, a) has infinite multiplicity.

A metric on the space of persistence diagrams is the bottleneck distance which bounds the Hausdorff distance [7]. It is given by

$$d_B(D(\mu), D(\nu)) = \inf_{\gamma} \sup_{p \in D(\mu)} \|p - \gamma(p)\|_\infty, \tag{3}$$

where the infimum is taken over all bijections $\gamma : D(\mu) \to D(\nu)$.

In [7], the following result is proven:

$$d_B(D(\mu), D(\nu)) \leq \|\mu - \nu\|_\infty \tag{4}$$

where $\mu, \nu : \mathbb{M} \to \mathbb{R}$ are tame functions. As an immediate consequence of (4), we can apply it to the model (1). Let $\Lambda_t(\beta, L)$ denote the subset of tame functions in $\Lambda(\beta, L)$ the class of Hölder functions

$$\Lambda(\beta, L) = \{f : S^2 \to \mathbb{R} \mid |f(x) - f(z)| \leq L(\arccos(x'y))^\beta, x, z \in S^2\}, \tag{5}$$

where $0 < \beta \leq 1$ and $L > 0$.

If we assume $\mu \in \Lambda_t(\beta, L)$ for the model (1) ϵ is $N(0, \sigma^2)$, for the estimator $\widehat{\mu}_n$ with $0 < \beta \leq 1$ and $L > 0$,

$$\sup_{\mu \in \Lambda_t(\beta, L)} \mathbb{E} d_B\left(D(\hat{\mu}_n), D(\mu)\right) \leq L^{2/(2\beta+2)} \left(\frac{\sigma^2 \ (\beta+2)2^3}{\beta^2} \ \frac{\log n}{n}\right)^{\beta/(2\beta+2)}$$

as $n \to \infty$ [3].

Consequently, because of the large n in medical image data, Figure 4 is an accurate description of the population parameters and therefore the extra layer of homology classes for the autistic group is most likely significant.

5 Discussion

We have presented the concept of persistence diagrams and described the filtration based algorithm for constructing the persistence diagrams. Since cortical thickness is highly noisy, kernel smoothing is applied to remove high frequency spatial noise before the filtration. The constructed degree 0 and 1 persistence diagrams seem to show a significant extra layer of homology classes for the autistic group (Figure 4). Cortical thickness has been shown to be a discriminating anatomical feature for autism although the underlying biological mechanism is still unknown [6]. It is found that autistic subjects exhibit more structural variability than normal controls. It is likely that autistic subjects exhibit a more complex underlying geometric and topological pattern of the brain which is the cause of additional anatomical variability.

It is unclear how one determines the possible statistical significance level of a seemingly significant extra layer of persistent homology classes. One may be tempted to use hypothesis-free classification frameworks for inference. However, Figure 4 shows that classification based on possibly discriminating spatial pattern might be difficult if not impossible. Note that the autistic scatter plots basically encompass the control scatter plots for the degree 0 and degree 1 persistence diagrams. Since there are a lot of overlap, it may be difficult to directly apply machine learning techniques.

It is hoped that this paper offers a spring board for further investigation of the persistence diagram based characterization of medical images. There are a lot of methodological issues we have not discussed in the paper such as an inferential procedure or the estimation of confidence circles around paired points possibly via the bootstrap. These we hope to do in future works.

Acknowledgment

Authors wish to thank Vikas Singh of the University of Wisconsin-Madison for discussion on the persistence diagram.

References

1. S. Angenent, S. Hacker, A. Tannenbaum, and R. Kikinis. On the laplace-beltrami operator and brain surface flattening. *IEEE Transactions on Medical Imaging*, 18:700–711, 1999.
2. C. Brechbuhler, G. Gerig, and O. Kubler. Parametrization of closed surfaces for 3d shape description. *Computer Vision and Image Understanding*, 61:154–170, 1995.
3. P. Bubenik, G. Carlsson, P.T. Kim, and Z. Luo. Asymptotic minimax sup-norm risk on manifolds with application to topology. *Preprint*, 2008.
4. P. Bubenik and P.T. Kim. Aa statistical approach to persistent homology. *Homology Homotopy and Applications*, 9:337–362, 2007.
5. M.K. Chung, L. Shen Dalton, K.M., A.C. Evans, and R.J. Davidson. Weighted fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566–581, 2007.
6. M.K. Chung, S. Robbins, Davidson R.J. Alexander A.L. Dalton, K.M., and A.C. Evans. Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, 25:1256–1265, 2005.
7. D. Cohen-Steiner, H. Edelsbrunner, and J. Harer. Stability of persistence diagrams. *Discrete and Computational Geometry*, 37, 2007.
8. D.L. Collins, P. Neelin, T.M. Peters, and A.C. Evans. Automatic 3d intersubject registration of mr volumetric data in standardized talairach space. *J. Comput. Assisted Tomogr.*, 18:192–205, 1994.
9. Ghrist R. de Silva, V. Homological sensor networks. *Notic Amer Math Soc*, 54:10–17, 2007.
10. V. de Silva and P. Perry. Plex version 2.5. *http://math.stanford.edu/comptop/programs/plex*, 2008.
11. Dequent M-L. Mileyko Y. Edelsbrunner, H. and O. Pourquie. Assessing periodicity in gene expression as measured by microarray data. *Preprint*, 2008.
12. H. Edelsbrunner and J. Harer. Persistent homology - a survey. In *Twenty Years After, American Mathematical Society*, page in press, 2008.
13. H. Edelsbrunner, D. Letscher, and A. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.
14. B. Fischl and A.M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *PNAS*, 97:11050–11055, 2000.
15. K.J. Friston. A short history of statistical parametric mapping in functional neuroimaging. Technical Report Technical report, Wellcome Department of Imaging Neuroscience, ION, UCL., London, UK., 2002.
16. X. Gu, Y.L. Wang, T.F. Chan, T.M. Thompson, and S.T. Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging*, 23:1–10, 2004.
17. M. K. Hurdal and K. Stephenson. Cortical cartography using the discrete conformal approach of circle packings. *NeuroImage*, 23:S119S128, 2004.
18. J. O. Irwin. The frequency distribution of the difference between two independent variates following the same poisson distribution. *Journal of the Royal Statistical Society: Series A*, 100:415416, 1937.

19. S.C. Joshi. Large deformation diffeomorphisms and gaussian random fields for statistical characterization of brain sub-manifolds. 1998.
20. S.J. Kiebel, J.-P. Poline, K.J. Friston, A.P. Holmes, and K.J. Worsley. Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. *NeuroImage*, 10:756–766, 1999.
21. J. Klemelä. Asymptotic minimax risk for the white noise model on the sphere. *Scand J Statist*, 26:465–473, 1999.
22. K. Kollakian. Performance analysis of automatic techniques for tissue classification in magnetic resonance images of the human brain. Technical Report Master's thesis, Concordia University, Montreal, Quebec, Canada, 1996.
23. J. P. Lerch and A.C. Evans. Cortical thickness analysis examined through power analysis and a population simulation. *NeuroImage*, 24:163–173, 2005.
24. E. Luders, K.L. Narr, P.M. Thompson, D.E. Rex, R.P. Woods, and Jancke L. andToga A.W. DeLuca, H. Gender effects on cortical thickness and the influence of scaling. *Human Brain Mapping*, 27:314–324, 2006.
25. J.D. MacDonald, N. Kabani, D. Avis, and A.C. Evans. Automated 3-D extraction of inner and outer surfaces of cerebral cortex from mri. *NeuroImage*, 12:340–356, 2000.
26. M.I. Miller, A. Banerjee, G.E. Christensen, S.C. Joshi, N. Khaneja, U. Grenander, and L. Matejic. Statistical methods in computational anatomy. *Statistical Methods in Medical Research*, 6:267–299, 1997.
27. M.I. Miller, A.B. Massie, J.T. Ratnanather, K.N. Botteron, and J.G. Csernansky. Bayesian construction of geometrically based cortical thickness metrics. *NeuroImage*, 12:676–687, 2000.
28. J. Milnor. *Morse Theory*. Princeton University Press, 1973.
29. D. Morozov. *Homological Illusions of Persistence and Stability*. Ph.D. thesis, Duke University, 2008.
30. D.Q. Naiman. Volumes for tubular neighborhoods of spherical polyhedra and statistical inference. *Ann. Statist.*, 18:685–716, 1990.
31. T. Nichols and S. Hayasaka. Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med. Res.*, 12:419–446, 2003.
32. Ozturk O. Ferhatosmanoglu H. Sacan, A. and Y. Wang.
33. D.O. Siegmund and K.J. Worsley. Testing for a signal with unknown location and scale in a stationary gaussian random field. *Annals of Statistics*, 23:608–639, 1996.
34. J.G. Sled, A.P. Zijdenbos, and A.C. Evans. A nonparametric method for automatic correction of intensity nonuniformity in mri data. *IEEE Transactions on Medical Imaging*, 17:87–97, 1988.
35. J.E. Taylor and K.J. Worsley. Random fields of multivariate test statistics, with applications to shape analysis. *Annals of Statistics*, page in press, 2008.
36. B. Timsari and R. Leahy. An optimization method for creating semi-isometric flat maps of the cerebral cortex. In *The Proceedings of SPIE, Medical Imaging*, 2000.
37. K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.
38. A. Yezzi and J.L. Prince. An eulerian pde approach for computing tissue thickness. *IEEE Transactions on Medical Imaging*, 22:1332–1339, 2003.
39. A.J. Zomorodian. *Computing and Comprehending Topology: Persistence and Hierarchical Morse Complexes*. Ph.D. Thesis, University of Illinois, Urbana-Champaign, 2001.
40. A.J. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.

Part II. Topological Characterization of Signal in Brain Images Using Min-max Diagrams

Moo K. Chung[12], Vikas Singh[1], Peter T. Kim[4],
Kim M. Dalton[2], Richard J. Davidson[23]

[1]Department of Biostatistics and Medical Informatics
[2]Waisman Laboratory for Brain Imaging and Behavior
[3]Department of Psychology and Psychiatry
University of Wisconsin, Madison, WI 53706, USA
[4] Department of Mathematics and Statistics
University of Guelph, Guelph, Ontario N1G 2W1, Canada

mkchung@wisc.edu

Abstract. We present a novel computational framework for characterizing signal in brain images *via* nonlinear pairing of critical values of the signal. Among the astronomically large number of different pairings possible, we show that representations derived from specific pairing schemes provide concise representations of the image. This procedure yields a "min-max diagram" of the image data. The representation turns out to be especially powerful in discriminating image scans obtained from different clinical populations, and directly opens the door to applications in a variety of learning and inference problems in biomedical imaging. It is noticed that this strategy significantly departs from the standard image analysis paradigm – where the 'mean' signal is used to characterize an ensemble of images. This offers robustness to noise in subsequent statistical analyses, for example; however, the attenuation of the signal content due to averaging makes it rather difficult to identify subtle variations (which may be clinically relevant). The proposed topologically oriented method seeks to address these limitations by characterizing and encoding *topological* features or attributes of the image. As an application, we have used this method to characterize (in the form of min-max diagrams) cortical thickness measures along brain surfaces in classifying autistic subjects. Our promising experimental results provide evidence of the power of this representation.

1 Introduction

The use of critical values of measurements (e.g., in images) within classical image analysis and computer vision has been relatively limited so far, and typically appear as part of simple *preprocessing* tasks such as feature extraction and identification of "edge pixels" in an image. For example, first or second order image derivatives may be used to identify the edges of objects (e.g., LoG mask) to serve as the contour of an anatomical shape, possibly using priors to provide

additional shape context. Specific properties of critical values as a topic on its own, however, has received less attention. Part of the reason is that it is difficult to construct a streamlined linear analysis framework using critical points, or values of images. Also, the computation of critical values is a nonlinear process and almost always requires the numerical estimation of derivatives. In some applications where this is necessary (e.g., in level sets [8]), the discretization scheme must be chosen carefully, and remains an active area of research [8]. It is noticed that in most of these applications, the interest is only in the stable estimation of these points rather than (1) their *properties*, and (2) how these properties vary as a function of images. We note that in brain imaging, on the other hand, the *use* of extreme values has been quite popular in other types of problems. For example, these ideas are employed in the context of correction for multiple comparisons using random field theory [11]. Recall that in random field theory, the extreme of a statistic is obtained from an ensemble of images, and is used to compute the p-value (for correcting for correlated noise across neighboring voxels). Our interest in this paper is to take a topologically oriented view of the image data. We seek to interpret the critical values in this context and assess their response as a function of brain image data. In particular, we explore specific representation schemes (to be introduced shortly) and evaluate the benefits they afford with respect to different applications.

The calculation of the critical values of a certain function of images (e.g., image intensities, cortical thickness, curvature maps etc.) is the first step of our procedure. This is performed after heat kernel smoothing [3], which allows us to obtain analytic estimation of the derivatives. It is the second step which is more interesting, and a central focus of the paper. The obtained critical values are paired in a nonlinear fashion following a specific pairing rule to produce so-called min-max diagrams. These are similar to the theoretical construct of persistence diagrams in algebraic topology and computational geometry, but have notable differences (discussed in §2.2). Min-max diagrams resemble scatter plots, and lead to a powerful representation of the key characteristics of their corresponding images. We discuss these issues in detail, and provide a number of examples and experiments to highlight their key advantages, limitations, and possible applications to a wide variety of medical imaging problems.

This paper makes the following contributions: (1) We propose a new topologically oriented data representation framework using the min-max diagrams; (2) We present a new $\mathcal{O}(n \log n)$ algorithm for generating such diagrams without having to modify or adapt the complicated machinery used for constructing persistence diagrams [6]; (3) Using brain MRI, we demonstrate that using the min-max diagram representation, upon choice of a suitable kernel function, the subsequent classification task (e.g., using support vector machines) becomes very simple. In other words, because this representation captures the relevant features of the image nicely, it induces separability in the distribution of clinically different populations (e.g., diseased and control groups). We show that significant improvements can be obtained over existing techniques.

2 Main Ideas

Consider measurements f from images given as

$$f(t) = \mu(t) + \epsilon(t),\ t \in \mathbb{M} \subset \mathbb{R}^d, \tag{1}$$

where μ is the unknown mean signal (to be estimated) and ϵ is noise. The unknown mean signal is estimated *via* image smoothing over $\mathbb{M}$, and denoted as $\widehat{\mu}$. Traditionally, the estimate for the residual $f - \widehat{\mu}$ is used to construct a test statistic corresponding to a hypothesis about the signal. The mean signal may not be able to fully characterize complex imaging data, and as a result, may have limitations in the context of inference. Hence, we propose to use a new topologically motivated framework called the *min-max diagram*, which is the scatter plot of specific pairing of critical values. Intuitively, the collection of critical values of μ can approximately characterize the shape of the continuous signal μ. By pairing critical values in a nonlinear fashion and plotting them, we construct the min-max diagram. We will provide additional details shortly.

2.1 Heat Kernel Smoothing

In order to generate the min-max diagram, we need to find the critical values of μ. It requires estimating the unknown signal smoothly so that derivatives can be computed. We avoid the diffusion equation based implicit smoothing techniques [2] since the image derivatives can only be obtained *via* finite difference type approaches. Instead, we present a more flexible spectral approach called *heat kernel smoothing* that explicitly represents the solution to the diffusion equation analytically. Heat kernel smoothing analytically solves the following equation

$$\frac{\partial F}{\partial \sigma} = \Delta F,\ F(t, \sigma = 0) = f(t).$$

The solution is given in terms of eigenfunctions ψ_k (and the corresponding eigenvalues λ_k) of the Laplace-Beltrami operator, i.e., $\Delta f + \lambda f = 0$. Define the heat kernel K_σ as

$$K_\sigma(t, s) = \sum_{k=0}^{\infty} e^{-\lambda_k \sigma} \psi_k(t) \psi_k(s).$$

The heat kernel smoothing estimate of μ is then given by

$$\widehat{\mu} = \int_{\mathbb{M}} K_\sigma(t, s) f(s)\ d\eta(s) = \sum_{i=0}^{\infty} e^{-\lambda_k \sigma} f_k \psi_k(t). \tag{2}$$

The Fourier coefficients $f_k = \int_{\mathbb{M}} f \psi_k\ d\eta$ are estimated using least squares [10]. For a given bandwidth, the expansion is truncated at the degree where adding more terms will not increase the goodness of fit.

Examples. For $\mathbb{M} = [0, 1]$, with the additional constraints $f(t + 2) = f(t)$ and $f(t) = f(-t)$, the eigenfunctions are

$$\psi_0(t) = 1, \psi_k(t) = \sqrt{2} \cos(k\pi t)$$

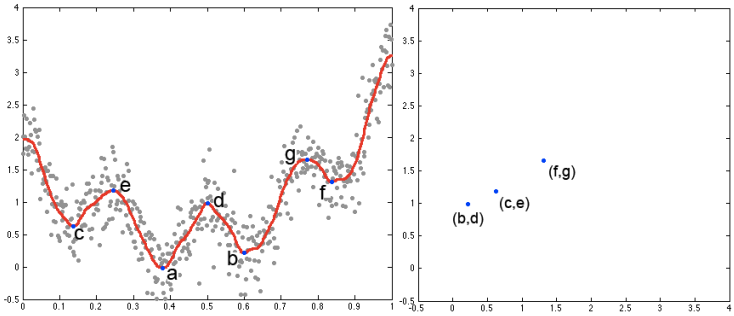

Fig. 1. The birth and death process of sublevel sets. Here $a < b < c < f$ are minimums and $d < e < g$ are maximums. At $y = b$, we add a new component to the sublevel set. When we increase the level to $y = d$, we have the death of the component so we pair them. In this simulation, we pair (f, g), (c, e) and (b, d) in the order of parings generated in Algorithm 1.

with the corresponding eigenvalues $\lambda_k = k^2\pi^2$. For simulation in Fig. 1, we used $\sigma = 0.0001$ and truncated the series at the 100-th degree.

For $\mathbb{M} = S^2$, the eigenfunctions are the spherical harmonics $Y_{lm}(\theta, \varphi)$ and the corresponding eigenvalues are $\lambda_l = l(l+1)$. The bandwidth $\sigma = 0.001$ and degree $k = 42$ was used for cortical thickness example in Fig. 2. We found that bandwidths larger than 0.001 smooth out relevant anatomical detail.

Analytic Derivatives. The explicit analytic derivative of the expansion (2) is simply given by

$$\mathcal{D}\widehat{\mu} = \sum_{i=0}^{\infty} e^{-\lambda_i \sigma} f_i \mathcal{D}\psi_i(t)$$

where $\mathcal{D}$ is $\frac{\partial}{\partial t}$ for $[0, 1]$ and $(\frac{\partial}{\partial \theta}, \frac{\partial}{\partial \varphi})$ for S^2. For the unit interval, the derivatives are $\mathcal{D}\psi_l(t) = -\sqrt{2}l\pi \sin(l\pi t)$. For S^2, the partial derivatives with respect to θ are given in iterative formulas. Fig. 2 shows the result of minimum and maximum detection after heat kernel smoothing.

2.2 Min-max Diagram

A function is called a Morse function if all critical values are distinct and non-degenerate, i.e., the Hessian does not vanish. For images (where intensities are given as integers), critical values of intensity may not all be distinct; however, the underlying continuous signal μ in (1) can be assumed to be a Morse function. Assuming $\widehat{\mu}$ is a Morse function, we define a sublevel set as $R(y) = \widehat{\mu}^{-1}(-\infty, y]$. The sublevel set is the subset of $\mathbb{M}$ that satisfies $\widehat{\mu}(t) \leq y$. As we increase y from $-\infty$, the number of connected components of $R(y)$ changes as we pass through critical values.

Let us denote the local minimums as $g_1, \cdots, g_m$ and the local maximums as $h_1, \cdots, h_n$. Since the critical values of a Morse function are all distinct, we can strictly order the local minimums from the smallest to the largest as $g_{(1)} < g_{(2)} < \cdots < g_{(m)}$ and similarly for the local maximums as $h_{(1)} < h_{(2)} < \cdots < h_{(n)}$ by sorting them. At each minimum, the sublevel set adds a new component while

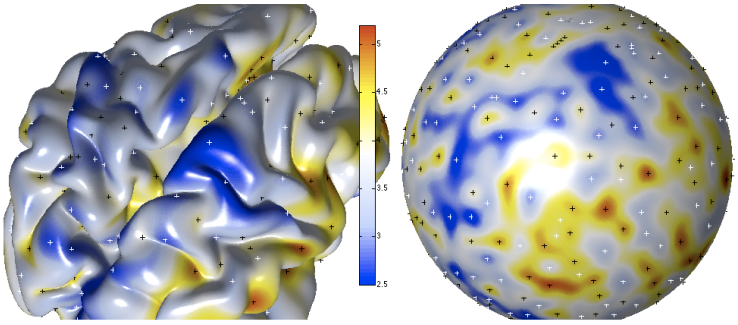

Fig. 2. Heat kernel smoothing of cortical thickness and surface coordinates with $\sigma = 0.001$ and degree $k = 42$. For better visualization, it has been flattened onto the unit sphere. The white (black) crosses are local minimums (maximums). They will be paired in a specific manner to obtain the min-max diagram. The min-max diagram is invariant to whether it is constructed from the cortical surface or from the unit sphere.

at a local maximum, two components merge into one. By keeping track of the birth and death of components, it is possible to compute topological invariants of sublevel sets such as Euler characteristics and Betti numbers (see [6]).

Simulation. The birth and death processes are illustrated in Fig. 1, where the gray dots are simulated with Gaussian noise with mean 0 and variance 0.2^2 as

$$f(t) = t + 7(t - 1/2)^2 + \cos(8\pi t)/2 + N(0, 0.2^2).$$

The signal is estimated and plotted as the red line using the 1D heat kernel smoothing in §2.1. Let us increase y from $-\infty$ to ∞. At $y = b$, we add a new component to the sublevel set $R(y)$. When we increase the level to $y = d$, we have the death of the component so we pair b and d. In this simulation, we need to pair (b, d), (c, e) and (f, g).

Pairing Rule. *When we pass a maximum and merge two components, we pair the maximum with the higher of the minimums of the two components* [6]. Doing so we are pairing the birth of a component to its death. Note that the paired critical values may not be adjacent to each other. The min-max diagram is then defined as the scatter plot of these pairings.

For higher dimensional Morse functions, saddle points can also create or merge sublevel sets so we also have to be concerned with them. If we include saddle points in the pairing rule, we obtain *persistence diagrams* [6] instead of min-max diagrams. In one dimension, the two diagrams are identical since there are no saddle points in 1D Morse functions. For higher dimensions, persistence diagrams will have more pairs than min-max diagrams. The addition of the saddle points makes the construction of the persistence diagrams much more complex and will be pursued in subsequent research. We note that [12] presents an algorithm for generating persistence diagrams based on filteration of Morse complexes.

Algorithm. We have developed a new simpler algorithm for pairing critical values. Our algorithm generates min-max diagrams as well as persistence diagrams

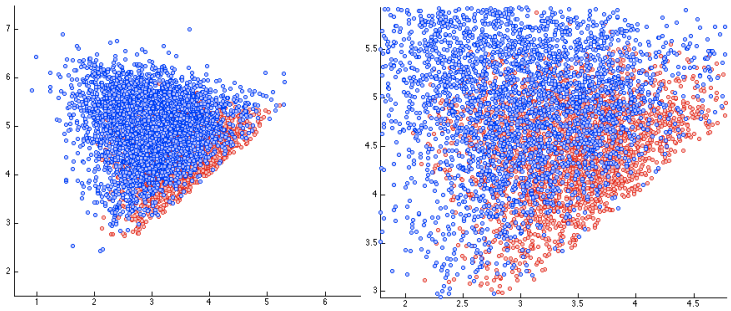

Fig. 3. Min-max diagram for 11 control (blue) and 16 autistic (red) subjects. The pairings for autism often occurs closer to $y = x$ line indicating there is greater high frequency noise in autism. This observation is consistent with the autism literature where it has been found that there is greater anatomical variability in autism subjects than the controls subjects. This figure suggests that the min-max diagram may indeed be useful for discriminating populations.

for 1D Morse functions. At first glance, the nonlinear nature of pairing does not seem to yield a straightforward algorithm. The trick is to start with the maximum of minimums and go down to the next largest minimum in an iterative fashion. The algorithm starts with $g_{(m)}$ (step 3). We only need to consider maximums above $g_{(m)}$ for pairing. We check if maximums h_j are in a neighborhood of $g_{(m)}$, i.e. $h_j \sim g_{(m)}$. The only possible scenario of not having any larger maximum is when the function is unimodal and obtains the global minimum $g_{(m)}$. In this situation we have to pair $(g_{(m)}, \infty)$. Since ∞ falls outside our 'plot', we leave out $g_{(m)}$ without pairing. Other than this special case, there exists at least one smallest maximum h_m^* in a neighborhood of $g_{(m)}$ (intuitively, if there is a valley, there must be mountains nearby). Once we paired them (step 4), we delete the pair from the set of extreme values (step 5) and go to the next maximum of minimums $g_{(m-1)}$ and proceed until we exhaust the set of all critical values (step 6). Due to the sorting of minimums and maximums, the running time is $\mathcal{O}(n \log n)$. This may also be implemented using a plane-sweep approach [4] which also gives a running time of $\mathcal{O}(n \log n)$. In this case, pairing will be based on how points enter or leave the queue of "events" as the plane (or line) sweeps in the vertical direction.

Algorithm 1 Iterative Pairing and Deletion

1. $H \leftarrow \{h_1, \cdots, h_n\}$.
2. $i \leftarrow m$.
3. $h_i^* = \arg\min_{h_j \in H}\{h_j | h_j > g_{(i)}, h_j \sim g_{(i)}\}$.
4. *If* $h_i^* \neq \emptyset$, *pair* $(g_{(i)}, h_i^*)$
5. $H \leftarrow H - h_i^*$.
6. *If* $i > 1$, $i \leftarrow i - 1$ *and go to Step 3.*

3 Experimental Results

We used an MRI dataset of 16 highly functional autistic subjects and 11 normal control subjects (aged-matched right-handed males). These images were ob-

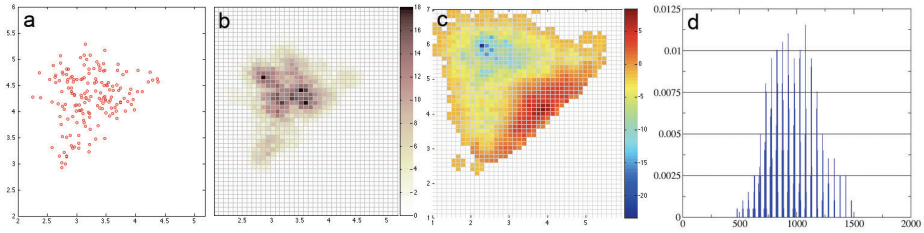

Fig. 4. (a) Min-max diagram of an autistic subject from Fig. 2. (b) The concentration map of the min-max diagram is constructed by discretizing the square $[1,7]^2$ into 50^2 uniform pixels and evaluating the number of pairs within a circle ($r = 0.2$) centered on the pixel. (c) The t-test statistic (autism - control) shows significant group differences in red regions ($t \geq 3.61$) *vs* blue ($t \leq -4.05$) regions at level 0.05 (corrected). (d) PDF of the concentration map.

tained from a 3-Tesla GE SIGNA scanner, and went through intensity nonuniformity correction, spatially normalized into the MNI stereotaxic space, and tissue segmentation. A deformable surface algorithm was used to obtain the inner cortical surface by deforming from a spherical mesh (see Fig. 2). The outer surface $\mathbb{M}$ was obtained by deforming the inner surface further. The cortical thickness f is then defined as the distance between the two surfaces [7], this measure is known to be relevant for autism. Since the critical values do not change even if we change the underlying manifold from $\mathbb{M}$ to S^2, the min-max diagram must be invariant as well. Therefore, the min-max diagram is constructed on the unit sphere by projecting the cortical data on to the sphere. Fig. 3 shows the superimposed min-max diagram for 11 control (blue) and 16 autistic (red) subjects, and a single subject example is shown in Fig. 4. Pairings for autistic subjects are more clustered near $y = x$ indicating higher frequency noise in autism. More pairing occurs at high and low thickness values in the control subjects showing additional topological structures not present in autism.

Statistical Inference. We have formally tested our hypothesis of different topological structures between the groups. Given a min-max diagram in the square $[1,7]^2$, we have discretized the square with the uniform grid such that there are a total of 50^2 pixels (see Fig. 4-b). A concentration map of the pairings was obtained by counting the number of pairs in a circle of radius 0.2 centered at each pixel. Notice that this approach is somewhat similar to the voxel-based morphometry [1], where brain tissue density maps are used as a shapeless metric for characterizing concentration of the amount of tissue. The inference at 0.05 level (corrected for multiple comparison) was done by performing 5000 random permutations on the maximum of t-statistic of concentration maps (Fig. 4-c).

SVM Based Classification. Our final set of experiments were performed to evaluate the usefulness of min-max diagrams for classification at the level of individual subjects. We view the concentration map of each min-max diagram as a PDF (Fig. 4), which allows easy construction of appropriate kernels and making use of Support Vector Machines (SVM). We evaluated linear and Gaussian weighted kernels (using Bhattacharya distance between the two PDFs [5]) and found that the accuracy results were quite similar. To perform our evaluations

relative to existing techniques, we used data shared with us by the authors in [9]. We summarize our results next.

For k-fold cross-validation, by varying $k \in \{9, \cdots, 2\}$, and performing 30 random runs for each k value (calculating the mean accuracy), we consistently achieved near perfect accuracy. The algorithm performs exceedingly well for 2-fold cross-validation as well – when only one half of the data is used as the training set. We incrementally decreased the size of the training set (up to 35%) and found that the algorithm still gives more than 96% accuracy. A simple comparison with $\sim 90\%$ accuracy estimates reported in [9] for this data suggest that the improvements in accuracy comes primarily from our min-max representation.

4 Conclusions

We have presented a unified framework of the min-max diagram based signal characterization in images. While unconventional, we believe that this representation is very powerful and holds considerable promise for a variety of learning and inference problems in neuroimaging. To demonstrate these ideas, we applied the methods to characterize cortical thickness data in a dataset of autistic and control subjects, *via* the use of a new Iterative Pairing and Deletion algorithm (to generate the min-max diagram). Our results indicate that significant improvements in classification accuracy are possible (relative to existing methods) merely by representing the input data as a set of min-max diagrams. Finally, we note that this paper only scratches the surface, and future research will clearly bring up other applications where these ideas might be useful.

References

1. J. Ashburner and K. Friston. Voxel-based morphometry - the methods. *NeuroImage*, 11:805–821, 2000.
2. A. Cachia, J.-F. Mangin, D. Riviere, et al. A primal sketch of the cortex mean curvature: a morphogenesis based approach to study the variability of the folding patterns. *IEEE Transactions on Medical Imaging*, 22:754–765, 2003.
3. M.K. Chung, L. Shen Dalton, K.M., A.C. Evans, and R.J. Davidson. Weighted fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging*, 26:566–581, 2007.
4. M. De Berg, O. Cheong, and M. van Kreveld. *Computational Geometry: Algorithms and Applications*. Springer, 2008.
5. E. Deza and M.M. Deza. *Dictionary of Distances*. Elsevier Science, 2006.
6. H. Edelsbrunner and J. Harer. Persistent homology - a survey. In *Twenty Years After, American Mathematical Society*, page in press, 2008.
7. B. Fischl and A.M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *PNAS*, 97:11050–11055, 2000.
8. S. Osher and R.P. Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Springer, 2003.
9. V. Singh, L. Mukherjee, and M. K. Chung. Cortical surface thickness as a classifier: Boosting for autism classification. In *Proc. of Medical Image Computing and Computer Assisted Intervention*, 2008.

10. M. Styner, I. Oguz, S. Xu, C. Brechbuhler, D. Pantazis, J. Levitt, M. Shenton, and G. Gerig. Framework for the statistical shape analysis of brain structures using spharm-pdm. In *Insight Journal, Special Edition on the Open Science Workshop at MICCAI*, pages 1–20, http:hdl.handle.net/1926/215, 2006.
11. K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.
12. A.J. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete and Computational Geometry*, 33:249–274, 2005.