

Topological Characterization of Signal in Brain Images Using Min-Max Diagrams

Moo K. Chung^{1,2}, Vikas Singh¹, Peter T. Kim⁴, Kim M. Dalton²,
and Richard J. Davidson^{2,3}

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. 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 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 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 the discretization scheme must be chosen carefully, and remains an active area of research. 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 multiple comparison correction using random field theory [9]. 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 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]. 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 [6] 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 [2] [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., autism vs. controls). We show that significant improvements can be obtained over existing techniques.

2 Main Ideas

Consider measurements f from images given as

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

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 $\hat{\mu}$. Traditionally, the estimate for the residual $f - \hat{\mu}$ 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 [1] since the approach tend to result in unstable derivative estimation. Instead, we present a more flexible spectral approach called *heat kernel smoothing* that explicitly represents the solution to the diffusion equation analytically [3]. 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

$$\hat{\mu} = \int_{\mathbb{M}} K_\sigma(t, s) f(s) d\eta(s) = \sum_{i=0}^{\infty} e^{-\lambda_i \sigma} f_i \psi_i(t). \quad (2)$$

Examples. For $\mathbb{M} = [0, 1]$, with the additional constraints $f(t + 2) = f(t)$ and $f(t) = f(-t)$, the eigenfunctions are $\psi_0(t) = 1, \psi_k(t) = \sqrt{2} \cos(k\pi t)$ 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.

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

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

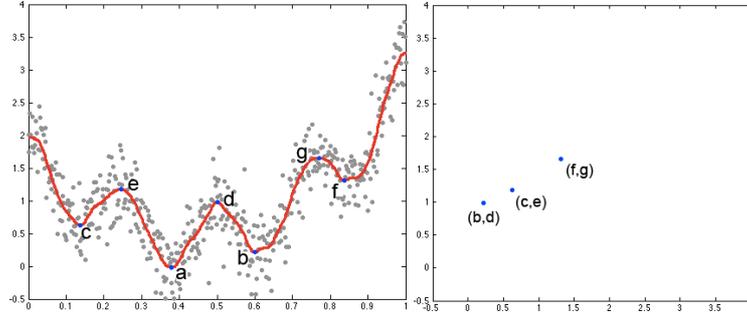


Fig. 1. The birth and death process of sublevel sets. Here $a < b < c < f$ are minimums and $d < e < g$ are maximums. At $y = b$, we add a new component to the sublevel set. When we increase the level to $y = d$, we have the death of the component so we pair them. In this simulation, we pair (f, g) ,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

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 can be given in slow iterative formulas. To speed up the computation through the paper, the convexity of the first order neighbor of a vertex in a cortical mesh is used in determining a critical point. 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. For a Morse function $\hat{\mu}$, define a sublevel set as $R(y) = \hat{\mu}^{-1}(-\infty, y]$. The sublevel set is the subset of \mathbb{M} satisfying $\hat{\mu}(t) \leq y$. 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, \dots, g_m and the local maximums as h_1, \dots, 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 $g_{(1)} < g_{(2)} < \dots < g_{(m)}$ and similarly for the local maximums as $h_{(1)} < h_{(2)} < \dots < h_{(n)}$ by sorting them. At each minimum, the sublevel set adds a new component while 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

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

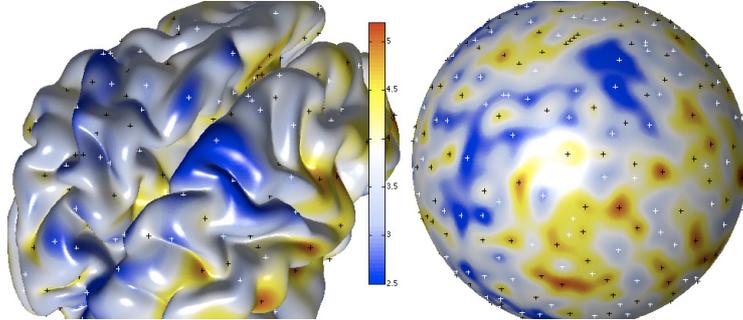


Fig. 2. Heat kernel smoothing of cortical thickness and surface coordinates with $\sigma = 0.001$ and degree $k = 42$. For better visualization, it has been flattened onto the unit sphere. The white (black) crosses are local minimums (maximums). They will be paired in a specific manner to obtain the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

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* [2] [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. We note that [10] presents an algorithm for generating persistence diagrams based on filtration 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 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

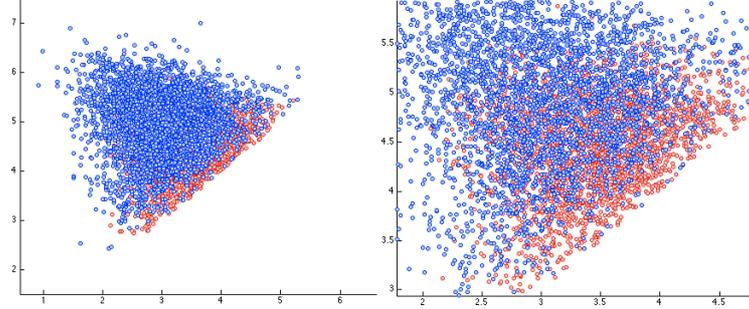


Fig. 3. Min-max diagram for 11 control (blue) and 16 autistic (red) subjects. The pairings for autism often occurs closer to $y = x$ line indicating there is greater high frequency noise in autism. This observation is consistent with the autism literature where it has been found that there is greater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

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. $H \leftarrow \{h_1, \dots, h_n\}$.
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.

Higher dimensional implementation is identical to the 1D version except how we define neighbors of a critical point. The neighborhood relationship \sim is established by constructing the Delaunay triangulation on all critical points.

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

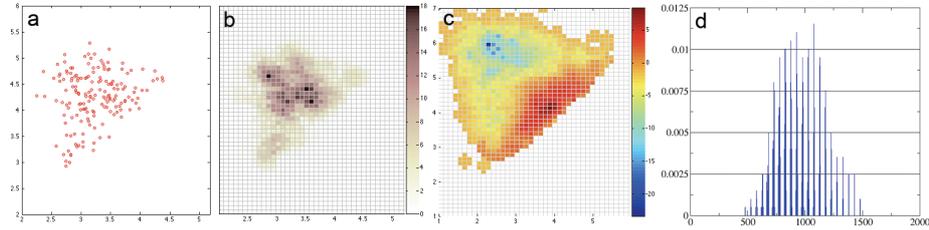


Fig. 4. (a) Min-max diagram of an autistic subject from Fig. 2. (b) The concentration map of the min-max diagram is constructed by discretizing the square $[1, 7]^2$ into 50^2 uniform pixels and evaluating the number of pairs within a circle ($r = 0.2$) centered on the pixel. (c) The t -test statistic (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.

obtained 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 [7] 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, this measure is known to be relevant for autism. Since the critical values do not change even if we geometrically change the underlying manifold from \mathbb{M} to S^2 , the min-max diagram must be topologically 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. 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 controls 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. 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).

If data is white noise, pairings occur close to $y = x$ line. The deviation from $y = x$ indicates signal. In the t -test result, we detected two main clusters of pairing difference. High number of pairings occurs around (2,6) for controls and (4,4) for autism. This is only possible if surfaces have more geometric features/signal in the controls. On the other hand, the autism shows noisier characteristic.

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 [8]. We summarize our results next.

For k -fold cross-validation, by varying $k \in \{9, \dots, 2\}$, 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 90% accuracy reported in [8] that uses the same data suggests 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. Cachia, A., Mangin, J.-F., Riviere, D., 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)
2. Chung, M.K., Bubenik, P., Kim, P.T.: Persistence diagrams of cortical surface data. In: *Proc. of Information Processing in Medical Imaging, IPMI* (2009)
3. Chung, M.K., Dalton, K.M., Shen, L., Evans, A.C., Davidson, R.J.: Weighted fourier representation and its application to quantifying the amount of gray matter. *IEEE Transactions on Medical Imaging* 26, 566–581 (2007)
4. De Berg, M., Cheong, O., van Kreveld, M.: *Computational Geometry: Algorithms and Applications*. Springer, Heidelberg (2008)
5. Deza, E., Deza, M.M.: *Dictionary of Distances*. Elsevier Science, Amsterdam (2006)

6. Edelsbrunner, H., Harer, J.: Persistent homology - a survey. A survey on Discrete and Computational Geometry: Twenty Years Later, 257–282 (2006)
7. MacDonald, J.D., Kabani, N., Avis, D., Evans, A.C.: Automated 3-D extraction of inner and outer surfaces of cerebral cortex from MRI. *NeuroImage* 12, 340–356 (2000)
8. Singh, V., Mukherjee, L., Chung, M.K.: Cortical surface thickness as a classifier: Boosting for autism classification. In: Proc. of Medical Image Computing and Computer Assisted Intervention (2008)
9. Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C.: A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping* 4, 58–73 (1996)
10. Zomorodian, A.J., Carlsson, G.: Computing persistent homology. *Discrete and Computational Geometry* 33, 249–274 (2005)