Cortical Surface-Based Analysis II: Inflation, Flattening, and a Surface-Based Coordinate System.

Bruce Fischl
Martin I. Sereno
and
Anders M. Dale

Abstract - The surface of the human cerebral cortex is a highly folded sheet with the majority of its surface area buried inside folds. As such, it is a difficult domain for computational as well as visualization purposes. We have therefore designed a set of procedures, which are an extension of previous work (Dale and Sereno 1993), for modifying the representation of the cortical surface to (i) inflate it so that activity buried inside sulci may be visualized, (ii) cut and flatten an entire hemisphere, and (iii) transform a hemisphere into a simple parameterizable surface such as a sphere for the purpose of establishing a surface-based coordinate system.

1. Introduction.

Currently, the most widely used method of analyzing functional brain imaging data is the projection of the functional data from a sequence of slices onto a standardized anatomical 3D space. The most common of these procedures is based on the Talairach atlas ((Talairach and Tournoux 1988), see e.g. (Collins, Neelin et al. 1994) for an automated procedure). While this type of approach has certain advantages (ease of use, widespread acceptance, applicability to subcortical structures), it also has significant drawbacks.

These drawbacks derive from the fact that the intrinsic topology of the cerebral cortex is that of a highly folded and curved 2D sheet. Estimates of the amount of “buried” cortex range from 60-70% (Zilles, Armstrong et al. 1988; Van Essen and Drury 1997) This implies that distances measured in 3D space will have an average error in the range of 45-50% with respect to the distance along the cortical sheet. In practice, errors can be much larger than this in cases where two points lie on different banks of a sulcus or gyrus.

From a functional standpoint, nonhuman primate neocortex is composed of a mosaic of visual, auditory, somatosensory, and motor areas, with visual areas alone occupying more than half of the total cortical surface (Felleman and Van Essen 1991; Kaas and Krubitzer 1991; Sereno and Allman 1991). The bulk of the remaining half is comprised of auditory, somatosensory,
motor, and limbic areas, each occupying about 1/8 of the total neocortex (Morel and Kaas 1992; Stepniewska, Preuss et al. 1993).

The majority of these areas are defined by their topographic maps of the sensory periphery (e.g. retinotopic, tonotopic, somatotopic). Typically, the metric encoding the relationship between these maps and the sensory periphery which they represent is not known (see, (Schwartz 1977; Schwartz 1980) for a notable exception). However, the two dimensional nature of the maps as well as their topographic arrangement strongly suggest that a two dimensional surface-based metric is more appropriate for analyzing their functional properties than the more typically used volume-based metrics.

The highly folded nature of the cortical surface also makes it difficult to view functional activity in a meaningful way. The typical means of visualization of this type of data is the projection of functional activation onto a set of orthogonal slices. This procedure is problematic as regions of activity which are close together in the volume may be relatively far apart in terms of the distance measured along the cortical surface. In addition, the naturally two-dimensional organization of cortical maps is largely obscured by the imposition of an external coordinate system in the form of orthogonal slices.

For these reasons we have developed a unified procedure which begins with a previously reconstructed cortex (Dale and Sereno 1993; Dale, Fischl et al. 1998) and modifies it in order to achieve three separate but related goals:

1. The “inflation” of the cortical surface so that activity occurring inside sulci may be easily visualized.

2. The flattening of an entire hemisphere so that the activity across the hemisphere may be seen from a single view, and so that computational procedures which are not tractable on arbitrary manifolds may be employed in the analysis of the cerebral cortex.

3. The “morphing” of a hemisphere into a surface, which maintains the topological structure\(^1\) of the original surface, but has a natural (i.e. closed-form) coordinate system.

Each of these procedures is accomplished in a manner that preserves as much of the topological and geometric structure of the original surface as possible. The methods described in this paper are an extension of previously presented work (Dale and Sereno 1993), and have been routinely used in a wide variety of studies (Sereno, Dale et al. 1995; Tootell, Reppas et al. 1995; Sereno, Dale et al. 1996; Tootell, Dale et al. 1996; Tootell, Dale et al. 1996; Tootell, Mendola et al. 1997).

\(^1\)The term *Topological structure* is frequently used to refer to the border of a domain as opposed to its global topology (Mortenson 1997). For example, once an incision has been made in the cortical surface it is topologically equivalent to a plane. Further incisions alter its topological structure, but not its topology (unless they result in multiple disconnected components).
2. Mapping of the cortical surface to parameterizable shapes.

Because of the varying intrinsic curvature of the cortical surface it is not possible to map it onto other significantly smoother surfaces (such as planes or spheres) without introducing some metric and/or topological distortion into the surface representations (Carmo 1976). A mapping between two surfaces with no metric distortion is called an isometry. Finding such a mapping from the sphere to the plane has been called the mapmaker's problem, and was shown to be impossible by Gauss in 1828 (Gauss 1828), as the surfaces in question have differing intrinsic (or Gaussian) curvature. Nevertheless, for the representations to be useful for either visualization or computational purposes, metric distortion must be minimized. Toward that end, we have developed a general procedure for minimizing metric distortion in a variety of contexts, such as surface inflation, flattening, as well as mapping to other parameterizable surfaces such as a sphere.

Constructing this type of mapping is a difficult task due to the complex and highly folded nature of the original surface, which requires a fine-scale tessellation in order to capture its metric and topological properties. One attractive means of flattening the surface is the method employed by Schwartz and colleagues (Schwartz, Shaw et al. 1989; Wolfson and Schwartz 1989; Schwartz 1990) in which the matrix of distances of each vertex to all other vertices is constructed in order to represent the metric properties of the original surface. The surface is then randomly projected onto a plane, and unfolded in such a way as to minimize the mean-squared error between the original distance matrix and that of the flattened surface. While this method is more than adequate for flattening small patches of the cortical surface, such as primary visual cortex to which it was originally applied, the computational requirements of the procedure in terms of both memory and time become prohibitive as the patch size grows.

A different type of method was employed by Dale and Sereno (Dale and Sereno 1993), and later by Carman (Carman, Drury et al. 1995) as well as Drury and colleagues ((Drury, Van Essen et al. 1996; Van Essen and Drury 1997; Van Essen, Drury et al. 1998). In this approach, a variety of local forces are constructed in order to encourage the preservation of local area and conformality (i.e. angle) while also forcing the surface to unfold onto a plane. These techniques have been successfully applied to entire cortical hemispheres, but suffer from a number of drawbacks. First, they require the use of terms such as a spring force in order to eliminate folds, which results in surfaces that are not optimal with respect to the preservation of any metric property. In addition, they treat the vertices on the borders of the flattened surface differently than those in the interior, thus constraining the shape of the resulting surface. Finally, they preserve local properties of the surface and therefore do not rule out large-scale distortions caused by locally correlated errors, although the use of multi-resolution techniques addresses this concern to some degree.

Part of the problem with applying the Schwartz method is that relatively long-range distances must be accounted for in order to unfold patches of cortex which have been folded by the projection process. They estimate that a procedure incorporating distance constraints on the order of 1 cm suffices to unfold monkey V1 (Schwartz, Shaw et al. 1989). Unfortunately, the distance required to smooth out a fold grows with the size of the surface (and the fold), quickly requiring untenable memory usage.
This problem occurs because distances are unoriented, and therefore mirror image configurations represent local minima in the energy functional. To see this, imagine a piece of paper, which is folded exactly along a string of vertices. If only nearest neighbor distances are being preserved, this represents an optimal configuration with the same energy as the completely unfolded state. The inclusion of neighborhoods which are small relative to the size of the entire sheet will not aid the problem, as the majority of the nodes on the surface are then beyond the neighborhood of the fold. This type of situation thus represents a local minimum, as moving vertices along the fold will increase the metric error until the rest of the surface expands. In order to cause the surface to unfold, a sufficient number of vertices must be included in the distance matrix so that the decrease in error caused by removing the fold more than offsets the increase in error of the region outside the fold, a solution that is not viable for as complex and large a surface as an entire cortical hemisphere.

We therefore construct a means of encouraging the surface to unfold which satisfies three criteria:

1) The final surface should be optimal with respect to minimizing metric distortion.

2) The borders of the cut surface should be treated no differently than the interior.

3) The resulting surface should have only minimal folding.

The first two criteria exclude the use of spring terms to “regularize” the mesh, which are typically introduced in order to prevent folding. Instead, we construct an energy functional that employs only a distance term for unfolded or positive regions of the surface, but applies an additional term to folded or negative regions\(^2\) in order to cause the surface to unfold.

### 2.1 Minimizing Metric Distortions.

The term that minimizes metric distortions is constructed as follows. Consider a mesh of \(V\) vertices distributed irregularly over a surface \(S\) embedded in a 3D Cartesian space. Denoting the distance between the \(i\)th and \(j\)th vertices at time \(t\) by \(d_{ij}^t\), we construct a mean-squared energy functional \(J_d:\)

\[
J_d = \frac{1}{2V} \sum_{i=1}^{V} \sum_{n \in N(i)} (d_{in}^t - d_{in}^0)^2,
\]

where \(x_i^t\) is the \((x,y,z)\) position of vertex \(i\) at time \(t\), \(d_{in}^0\) is the distance between the \(i\)th and \(n\)th vertices on the initial surface, and \(N(i)\) is the set of vertices defined to be in the neighborhood of vertex \(i\). Taking the gradient of \(J_d\) with respect to the \(k\)th vertex results in

---

\(^2\) We use the notion of an oriented area by defining a consistent normal direction on the surface (positive \(z\) in the plane, outward on a sphere). Any triangles in the tessellation in which the ordered cross-product of its legs is antiparallel to the normal direction is then assigned a negative area.
where $\mathbf{e}_{kn}$ is a unit vector pointing from vertex $k$ to vertex $n$.

### 2.2. Unfolding using oriented area.

As noted previously, causing the surface to unfold using only a distance term is not feasible for large surfaces. This is due to the fact that mirror-image configurations are not directly penalized, resulting in folded states that are local minima of the energy functional. These local minima are caused by the inherently unoriented nature of distances that do not explicitly distinguish between folded and unfolded states. In order to resolve this problem, we therefore seek an oriented metric property that discourages folds in the surface. The two obvious candidates are conformality and areal terms. While both can be employed successfully in this context, the use of an angle term results in a gradient which is dependent on the square of the inverse of the vertex spacing, and is therefore somewhat numerically unstable. In contrast, the use of an oriented area results in a quadratic energy functional with a well-defined minimum.

In order to define the areal term of the energy functional we consider the $i$th triangle in the surface tessellation depicted in Figure 1, with unit normal vector $\mathbf{n}_i$, and edges $\mathbf{a}_i$ and $\mathbf{b}_i$ connecting the vertex $\mathbf{x}_i$ to two of its neighbors (note that bold-faced symbols denote vector quantities). The unit normal $\mathbf{n}_i$ is given by the normalized cross product of the edges $\mathbf{a}_i$ and $\mathbf{b}_i$, while the area of the triangle is half the cross product of $\mathbf{a}_i$ and $\mathbf{b}_i$ dotted with the unit normal (i.e. the triple scalar product). If the normal vector field is given a consistent orientation on the surface\(^3\), then this becomes an oriented area, which may take on negative values indicating folds in the surface.

![Figure 1](image_url) Metric properties of the triangular tessellation.

Given this description of the metric properties of the surface through the triangular tessellation, we form an energy functional $J_a$ which penalizes negative area in proportion to the difference between the current area and the original area occupied by each triangle:

\[
\frac{\partial J_a}{\partial \mathbf{x}_k} = \frac{1}{V} \sum_{m \in N(k)} (d'_{kn} - d^0_{kn}) \mathbf{e}_{kn}
\]

---

\(^3\) This is always possible except in pathological cases such as the Möbius strip which are said to be nonorientable (Carmo 1976).
\[ J_a = \frac{1}{2T} \sum_{i=1}^{T} P(A'_i)(A'_i - A^0_i)^2, \quad P(A'_i) = \begin{cases} 1, & A'_i \leq 0 \\ 0, & \text{otherwise} \end{cases} \]

where, as before, superscripts denote time, with 0 being the initial areal values, \( T \) refers to the number of triangles in the tessellation, and the functional dependence of the \( A_i \)'s on the position of the vertex and its neighbors has been suppressed for simplicity of notation.

In order to minimize \( J_a \), we take the gradient with respect to the vertex positions \( \mathbf{x}_k \):

\[ \frac{\partial J_a}{\partial \mathbf{x}_k} = \frac{1}{T} \sum_{i=1}^{T} (A'_i - A^0_i) \frac{\partial A'_i}{\partial \mathbf{x}_k}. \]

Expanding the partial derivative using the chain rule yields:

\[ \frac{\partial A'_i}{\partial \mathbf{x}_k} = \frac{\partial A'_i}{\partial a_i} \frac{\partial a_i}{\partial \mathbf{x}_k} + \frac{\partial A'_i}{\partial b_i} \frac{\partial b_i}{\partial \mathbf{x}_k} \]

The partials of the change in the legs with respect to a change in the vertex position are dependent on what position the vertex in question occupies in a given triangle:

\[ \frac{\partial a_i}{\partial \mathbf{x}_k} = \begin{cases} [-1, -1, -1], & k = i \\ [1, 1, 1], & k = l \\ 0, & \text{otherwise} \end{cases} \quad \frac{\partial b_i}{\partial \mathbf{x}_k} = \begin{cases} [-1, -1, -1]^T, & k = i \\ [1, 1, 1], & k = l \\ 0, & \text{otherwise} \end{cases} \]

### 2.3. The complete energy functional.

The complete energy functional incorporating both distance and areal terms is given by

\[ J = \lambda_d J_d + \lambda_a J_a \]

where the \( \lambda_d \) and \( \lambda_a \) coefficients define the relative importance of unfolding versus the minimization of metric distortions respectively. Initially, \( \lambda_a \) takes on values much larger than \( \lambda_d \), and gradually decreases over time as the surface successfully unfolds. One additional point to note is that we smooth the gradients using iterative averaging during the numerical integration. This allows entire regions which are compressed or expanded to move coherently in the appropriate direction, and is similar to decimation followed by upsampling with interpolation. We allow each scale (defined by the number of iterations in the averaging) to equilibrate before reducing the scale and continuing. The actual minimization of \( J(\mathbf{x}) \) is accomplished using gradient descent with line minimization (Press, Teukolsky et al. 1994).

### 3. Surface Inflation.

The high degree of folding of the cortical surface makes it desirable to inflate the reconstructed surface for visualization purposes (Dale and Sereno 1993). This renders the interior of sulci visible, as well as making the surface-based distance between regions more apparent to visual inspection. The purpose of the surface inflation is thus to provide a representation of the cortical hemisphere that retains much of the shape and metric properties of the original surface, but allows the visualization of functional activity occurring within sulci. For this purpose, we define an energy function whose minimization results in the desired shape. This functional consists of two terms, a spring force which smooths the surface, and the metric-preservation term described in
section 2.1, which constrains the evolving surface to retain as much of the original metric properties as possible:

\[
J_s = \frac{1}{2V} \left( \sum_{i=1}^{V} \sum_{m \in N_1(i)} |x_i - x_m|^2 \right) + \lambda_s J_d,
\]

where \( N_1 \) denotes the set of nearest neighbors of each vertex, and \( J_d \) is as defined in section 2.1. We use Euler’s method with momentum to integrate \( J_s \) until the surface has achieved a desired smoothness as measured by the goodness-of-fit of the polyhedral approximation\(^4\).

![Figure 2. Inflated representations of the three cortical surfaces (sulci are dark and gyri are light).](image)

4. Flattening.

In order to flatten a cortical hemisphere with minimal distortion we make a number of cuts on the medial aspect of the original surface - one in a region around the corpus callosum to remove all subcortical structures, one down the fundus of the calcarine sulcus, and a set of equally spaced radial cuts. We then project the resulting surface onto a plane whose normal is given by the average surface normal of the cut surface. Once the projection has been accomplished, we again allow the surface to unfold by minimizing the energy given in section 2.3, using equally spaced randomly sampled distances in a 0.5 cm radius of each vertex as the neighborhood \( N(i) \). The result of this procedure is shown in Figure 3, which depicts three flattened left hemispheres.

![Figure 3. Three flattened left hemispheres (sulci are dark and gyri are light).](image)

\(^4\) We integrate the inflation functional until the normalized average distance of the neighbors of each vertex from its tangent plane is below a prespecified threshold.
5. **A surface-based coordinate system.**

The specification of corresponding points on different cortical surfaces requires the establishment of a uniform surface-based coordinate system. This is in contrast to volume-based coordinate systems in which a point on the cortical surface in one volume will typically not lie on the cortical surface of a different volume. In order to establish an inherently surface-based coordinate system, we morph the reconstructed cortical surface onto a parameterizable surface, as the parameterization then provides a natural coordinate system. The surface we choose for this purpose is a sphere for a number of reasons. Primarily, the mapping of the cortical hemisphere onto a sphere allows the preservation of the topological structure of the original surface (i.e. the local connectivity). This is in contrast to the use of a flattened surface, which requires cuts in order to lie flat with minimal distortion. These cuts change the topological structure of the surface, resulting in points on opposite sides of a cut, which are close to each other on the original cortical surface, becoming quite far apart on the final flattened representation. The choice of the sphere also allows us to retain much of the computational attractiveness of the, allowing the simple calculation of metric properties such as distances, areas and angles, properties that are more difficult to compute on more complex surfaces such as ellipsoids.

The process of unfolding the cortical surface on a sphere is identical to the procedure outlined in section 4, except that distances on the sphere are no longer Euclidean, but rather must be computed using the geodesics of the sphere. In addition, the lack of freedom to modify the shape of the unfolding surface necessitates the use of longer range distances than in the case of the flattening. Typically we minimize the metric distortion of a randomly spaced sampling of distances in a 1 cm radius of each vertex. Figure 4 illustrates the result of applying this procedure to three cortical hemispheres.

![Figure 4. Lateral view of three left hemispheres after morphing.](image)

6. **Conclusion.**

In this paper we have presented a unified set of procedures which transform the a previously reconstructed cortical surface, and are routinely used in our lab. These transformations
achieve two primary goals. First, they dramatically improve the ability to visualize functional activation taking place on the cortical surface. In addition, they allow two-dimensional analysis techniques to be applied to the functional and structural properties of the cortical surface. The mapping procedures we have presented have the advantage of being optimal with respect to a well defined energy functional that measures the amount of metric distortion of the transformed surface. In conjunction with segmentation methods (Dale and Sereno 1993; Atkins and Mackiewich 1996; Dale, Fischl et al. 1998; Teo, Sapiro et al. 1998), these procedures allow the routine use of surface-based representation and analysis for the first time.

7. References.


