Hypothesis Testing with Nonlinear Shape Models

Timothy B. Terriberry[1], Sarang C. Joshi[1,2], and Guido Gerig[1,3]

[1] Dept. of Computer Science
[2] Dept. of Radiation Oncology
[3] Dept. of Psychiatry
Univ. of North Carolina, Chapel Hill, NC 27599, USA
{tterribe,joshi,gerig}@cs.unc.edu

Abstract. We present a method for two-sample hypothesis testing for statistical shape analysis using nonlinear shape models. Our approach uses a true multivariate permutation test that is invariant to the scale of different model parameters and that explicitly accounts for the dependencies between variables. We apply our method to m-rep models of the lateral ventricles to examine the amount of shape variability in twins with different degrees of genetic similarity.

1 Introduction

We have been developing methods for statistical shape analysis utilizing medial representations. However, these and many other useful shape models contain a large number of parameters that lie in nonlinear spaces, and so traditional statistical analysis tools designed for Euclidean spaces have to be reformulated. In this paper we formalize the notion of hypothesis testing against data that lies in the direct product of a large number of nonlinear spaces as a tool for understanding growth and disease.

Recently, Fletcher et al. have developed methods for one-sample statistical shape analysis based on medial representations, or *m-reps* [1–3]. We turn to the problem of two-sample statistics, where we wish to answer the following question: given two samples from two different populations, do they have the same statistical distribution? This is the classic problem of testing the null hypothesis, H_0, that the populations are identical, against its complement, H_1. The main difficulty arises from the fact that m-reps lie on high-dimensional nonlinear manifolds where assumptions of Gaussianity are unreasonable, making traditional parametric or linear methods inapplicable.

We present a true multivariate permutation test approach that is equivalent to traditional nonparametric permutation tests in the univariate case, and converges to the same result as Hotelling's well-known T^2 test in the linear, normally-distributed case. The only tool we require on the underlying space our data lives in is the existence of a metric.

The mechanics of the method are similar to those used in correction for multiple tests [4]. Unlike methods of direct combination, which sum up various
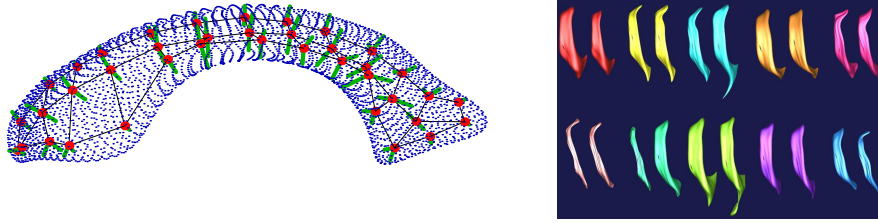
Fig. 1. Left: An example m-rep of a left lateral ventricle. The mesh vertices and off-shooting spokes make up the medial atoms. The shape the m-rep was fit to is shown as a point cloud surrounding it. Right: Ventricle pairs from five monozygotic twin pairs (top) and five dizygotic twin pairs (bottom).

test statistics [5, 6], our method is invariant to the scale of each term. This is critical when different shape parameters have different physical units and the choice of weighting between them can be arbitrary. Our test also accounts for the dependencies between model parameters.

1.1 A Metric Space for M-reps

M-reps are a medial shape model whose parameters provide intuitive descriptions of local object thickness, bending, narrowing, and widening. They have been well-described by previous authors [7], but for completeness we provide a brief summary. An m-rep is a coarse grid of samples that lie on the medial axis of an object. Each sample, called a *medial atom*, consists of a 4-tuple $\underline{m} = (\mathbf{x}, r, \mathbf{n}_0, \mathbf{n}_1)$ of parameters. The 3-D position of the atom is $\mathbf{x} \in \mathbb{R}^3$, the distance to the two closest boundary points is $r \in \mathbb{R}^+$, and $\mathbf{n}_0, \mathbf{n}_1 \in S^2$ are unit vectors that point from the atom position towards the two boundary points. The direct product of these spaces, $\mathbb{R}^3 \times \mathbb{R}^+ \times S^2 \times S^2$, is denoted $\mathcal{M}(1)$, and an entire m-rep with p medial atoms lives in the direct product space $\mathcal{M}(p) = \mathcal{M}(1)^p$. See Fig. 1 for an example of a complete model and a sample of our shape population.

Fletcher et al. treat medial atoms as elements of a Riemannian symmetric space [2]. Such a space is a differentiable manifold and has a Riemannian metric that is invariant to certain transformations of the space. $\mathbb{R}^3$ uses the normal Euclidean metric, while the positive reals, $\mathbb{R}^+$, use the metric $d(r_1, r_2) = |\log(r_1) - \log(r_2)|$, and the unit sphere, S^2, uses distance measured along the surface of the sphere. Every point on the manifold has a tangent plane, which is a vector space, and exponential and log maps that project from the plane to the manifold and back while preserving distances from the tangent point in a local neighborhood. For a more complete treatment, see Fletcher's Ph.D. thesis [3].

1.2 One-sample Statistics in Nonlinear Spaces

In linear spaces, the most important property of a probability distribution is often its first moment, the mean. Fréchet generalized the notion of an arithmetic mean of a sample of n points x_i drawn from a distribution in a general metric space M as the point which minimizes the sum-of-squared distances [8]:

$$\hat{\mu} = \text{argmin}_{x \in M} \frac{1}{2n} \sum_{i=1}^{n} d(x, x_i)^2 \ . \tag{1}$$

This is sometimes referred to as the Fréchet mean or the *intrinsic mean*, but hereafter will just be called the mean.

In general, this mean may not exist, or may not be unique, and without additional structure on the metric space, the minimization may be difficult to perform. However, for Riemannian manifolds, it is possible to compute the gradient of this functional [9], making a gradient descent algorithm possible [10]. Kendall showed that existence and uniqueness is guaranteed if the data is well-localized [11]. Fletcher et al. extend this, using principal component analysis (PCA) in the tangent plane at the mean to characterize the distribution of one sample [2].

1.3 Two-sample Statistics

If we assume both of our distributions are identical around the mean, and that they can be characterized entirely by the distance from the mean, then a single global distance value is sufficient to construct a univariate permutation test for equality of the two means. Permutation tests are appealing because they make no other distributional assumptions, requiring only that the data in each group be exchangeable under the null hypothesis that they do in fact come from the same distribution. The interested reader is referred to Bradley [12] or Nichols and Holmes [13] for details.

However, our geometric models contain parameters in nonlinear spaces, like the sphere. Some parameters may have a large variance, masking the effects of other variables with a smaller variance that might provide greater discrimination. Some may be highly correlated, unduly increasing their contribution to the distance over that of parameters with less correlation. Some will have completely different scales, and appropriate scale factors need to be determined to combine them in a single metric. These factors make the assumption that the distance from the mean entirely characterizes the distribution hard to justify.

For example, scaling the model will change the distance between medial atom centers, $\mathbf{x}$, without affecting the distance between radii or spoke directions. To combat this, Fletcher et al. propose scaling the latter by the average radius across corresponding medial atoms [2], but this choice is somewhat arbitrary. It does restore invariance to scale, but does nothing to handle differing degrees of variability or correlation. Different choices of scale factors will produce tests with different powers.

In $\mathbb{R}^n$, if we relax our assumption that the distribution is characterized by the distance from the mean, and instead assume only a common covariance, the classic Hotelling's T^2 test provides a test invariant to coordinate transformations. For normally distributed data, it is uniformly the most powerful (see a standard text, such as Anderson's [14], for a derivation). The test is based on the statistic: $T^2 \propto D^2 = (\hat{\mu}_1 - \hat{\mu}_2)^T \hat{\Sigma}^{-1} (\hat{\mu}_1 - \hat{\mu}_2)$, where $\hat{\mu}_1$ and $\hat{\mu}^2$ are the sample means and $\hat{\Sigma}$ the pooled sample covariance. Any linear change of coordinates yields a corresponding change in metric, but this is absorbed by the $\hat{\Sigma}^{-1}$ term.

2 Multivariate Permutation Tests

The hypothesis test we propose is an attempt to generalize the desirable properties of Hotelling's T^2 test to a nonparametric, nonlinear setting. We cannot take advantage of the vector space structure of the tangent plane, as Fletcher et al. do, to apply Hotelling's test directly, because there is a different tangent space around each sample's mean, and there may be no unique map between them. For example, on the sphere, such a map has one degree of freedom, allowing an arbitrary rotation of the coordinate axes in the vector space. Instead, we take a more general approach, only requiring that our objects lie in a metric space.

Our approach is based upon a general framework for nonparametric combination introduced by Pesarin [15]. The general idea is to perform a set of partial tests, each on a different aspect of the data, and then combine them into a single summary statistic, taking into account the dependence between the variables and the true multivariate nature of the data. We assume that we have two distributions with the same structure around the mean, and develop a test to determine if the means are equal. We now begin describing the details.

2.1 The Univariate Case

We begin by introducing notation and describing the procedure for a single, univariate permutation test. Suppose we have two data sets of size n_1 and n_2, $x_1 = \{x_{1,i}, i \in 1 \ldots n_1\}$ and $x_2 = \{x_{2,i}, i \in 1 \ldots n_2\}$, and a test statistic, $T(x_1, x_2)$. To test for a difference in the means, a natural test statistic is

$$T(x_1, x_2) = d(\hat{\mu}_1, \hat{\mu}_2) \,, \tag{2}$$

where $\hat{\mu}_1$ and $\hat{\mu}_2$ are the sample means of the two data sets computed via the optimization in (1). For other tests, other statistics are possible.

Under the null hypothesis, both samples are drawn from the same distribution, and so we may randomly permute the data between the two groups without affecting the distribution of $T(x_1, x_2)$. We pool the data together, and then generate $N = \binom{n_1 + n_2}{n_1}$ random partitions into two new groups, still of size n_1 and n_2. We label these $x_{1,i}^k$ and $x_{2,i}^k$, with $k \in 1 \ldots N$, and compute the value of the test statistic, T^k, for all of them. We always include the actual observed groupings

among this list, and denote its test statistic T^o. This forms an empirical distribution of the statistic, from which we can calculate the probability of observing T^o under the null hypothesis:

$$p(T^o) = \frac{1}{N} \sum_{k=1}^{N} H(T^k, T^o) \ , \qquad H(T^k, T^o) = \begin{cases} 1, & T^k \geq T^o \\ 0, & T^k < T^o \end{cases} . \qquad (3)$$

2.2 Partial Tests

If our data can be adequately summarized by a single test statistic, then this is the end of the story. We now turn to the case where we have M test statistics: one for each of the parameters in our shape model. Let $\mu_{1,j}$ and $\mu_{2,j}$ be the means of the jth model parameter for each population. Then we wish to test whether any hypothesis $H_{1,j} : \{\mu_{1,j} \neq \mu_{2,j}\}$ is true against the alternative, that each null hypothesis $H_{0,j} : \{\mu_{1,j} = \mu_{2,j}\}$ is true. The partial test statistics $T_j(x_1, x_2), j \in 1 \ldots M$ are defined analogously to (2), and the values for permutations of this data are denoted T_j^k, , with $j \in 1 \ldots M$, $k \in 1 \ldots N$.

Given that each $T_j(x_1, x_2)$ is significant for large values, consistent, and marginally unbiased, Pesarin shows that a suitable combining function (described in the next section) will produce an unbiased test for the global hypothesis H_0 against H_1 [15]. The meaning of each of these criteria is as follows:

1. **Significant for large values:** Given a significance level α and the critical value of $T_j(x_1, x_2)$ at α—T_j^α—the probability that $T_j^o \geq T_j^\alpha$ is at least α. For a two-sided test $T_j(x_1, x_2)$ must be significant for both large and small values.
2. **Consistent:** As the sample size $n = n_1 + n_2$ goes to infinity, the probability that $T_j^o \geq T_j^\alpha$ must converge to 1.
3. **Marginally unbiased:** For any threshold z, the probability that $T_j^o \leq z$ given $H_{0,j}$ must be greater than the probability that $T_j^o \leq z$ given $H_{1,j}$, irrespective of the results of any other partial test. This implies that T_j^o is positively dependent in $H_{1,j}$ regardless of any dependencies between variables.

Since each of our tests are restricted to the data from a single component of the direct product and we have assumed that the distributions around the means are identical, they are marginally unbiased. We cannot add a test for equality of the distributions about the mean, as then the test for equality of means would be biased on its outcome.

To illustrate these ideas, we present a simple example, which we will follow through the next few sections. We take two samples of $n_1 = n_2 = 10$ data points from the two-dimensional space $\mathbb{R} \times \mathbb{R}^+$, corresponding to a position and a scale parameter. The samples are taken from a multivariate normal distribution by exponentiating the second coordinate, and then scaling both coordinates by a factor of ten. They are plotted together in Fig. 2a. They have the common covariance (before the exponentiation) of $\frac{1}{2} \left(\begin{smallmatrix} 3 & 1 \\ 1 & 3 \end{smallmatrix} \right)$, and the two means are slightly offset in the second coordinate. That is, $\mu_{1,1} = \mu_{2,1}$, but $\mu_{1,2} < \mu_{2,2}$.
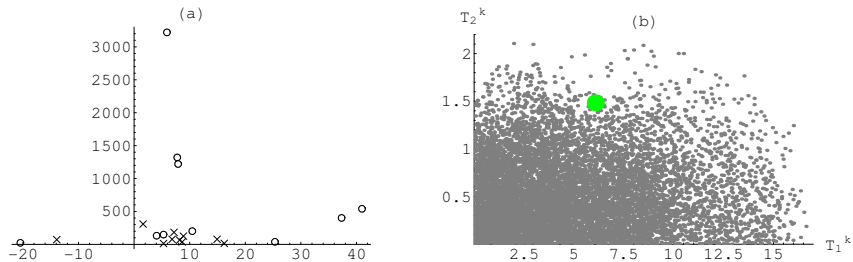
Fig. 2. The observed data and test statistics for our simple example. (a) shows the distribution of our two samples, with ×'s for the first and ∘'s for the second. (b) shows the distribution of the partial test statistics under permutation. The large dot indicates the location of the observed data point.

We construct $M = 2$ partial test statistics using (2) for each coordinate, and evaluate them using Monte Carlo simulation. To avoid an exponential complexity, we use a fixed $N = 10,000$ permutations, which still provides an unbiased test. The results are shown in Fig. 2b. The first partial test value lies in the middle of the distribution, while the second lies near the edge. However, the scale of the first test is much larger, because no logarithm is involved in its metric.

2.3 Multivariate Combination

Given the partial tests from the previous section, we wish to combine them into a single test, while preserving the underlying dependence relations between the tests. This is done in the following manner. We apply the same N permutations to the data when computing each of partial tests, and then compute a p-value using the empirical distribution for that test over all of the other permutations:

$$p(T_j^k) = \frac{1}{N} \sum_{l=1}^{N} H(T_j^l, T_j^k) \ . \tag{4}$$

Thus, for every permutation k we have a column vector of p-values, $p(T^k) = (p(T_1^k), \ldots, p(T_j^k), \ldots, p(T_M^k))^T$. It is critical to use the same permutations for each partial test, as this is what captures the nature of the joint distribution.

We now wish to design a combining function to produce a single summary statistic, $T'(p(T^k))$, from each p-value vector. For one-sided tests, this statistic must be monotonically non-increasing in each argument, must obtain its (possibly infinite) supremum when any p-value is zero, and the critical value T'^α must be finite and strictly smaller than the supremum. If these conditions are satisfied, along with those on the partial tests from the previous section, then $T'(p(T^k))$ will be an unbiased test for the global hypothesis H_0 against H_1 [15].

Our combining function is motivated by the two-sided case, where we can use the Mahalanobis distance. First, we compute a U^k vector for each permutation,

where $U_j^k = \Phi^{-1}(p(T_j^k) - \frac{1}{2N})$ and $j \in 1 \ldots M$. Here Φ is the cumulative distribution function for the standard normal distribution. The extra $\frac{1}{2N}$ term keeps the values finite when the p-value is 1, and is negligible as N goes to infinity.

Because the distribution of p-values for each partial test is uniform by construction, the marginal distribution of the U_j^k values over k for a single j is standard normal. Arranging these vectors into a single $N \times M$ matrix U, we can estimate the covariance matrix $\hat{\Sigma}_U = \frac{1}{N} U^T U$, and use the Mahalanobis statistic: $T'^k = (U^k)^T \hat{\Sigma}_U^{-1} U^k$. In the event that the data really is linear and normally distributed, $\hat{\Sigma}_U$ matrix converges to the true covariance as the sample size goes to infinity [16], making it asymptotically equivalent to Hotelling's T^2 test. Even if the sample size is small, the matrix Σ_U is well-conditioned regardless of the number of variables, since it is the covariance over the N permutations.

Typically, our distances are not signed, and so we are interested in a one-sided test. In this case, we use the positive half of the standard normal c.d.f., $U_j^k = \Phi^{-1}(1 - \frac{1}{2}(p(T_j^k) - \frac{1}{2N}))$, and assume the U^k distribution is symmetric about the origin. This assumption, however, implies that the covariance between $U_{j_1}^k$ and $U_{j_2}^k$ when $j_1 \neq j_2$ is exactly zero. The diagonal entries of $\hat{\Sigma}_U$ are 1 by construction, and so $\hat{\Sigma}_U = I$, the identity matrix. The fact that the p-values of the partial tests are invariant to scale obviates the need for arbitrary scaling factors. Thus, our one-sided combining function is:

$$T'^k = (U^k)^T \cdot U^k .\tag{5}$$

Note that normality of the partial test statistics is not required, and that the even though the marginal distributions of the U^k vectors are normal, the joint distribution may not be. Therefore, we must use a nonparametric approach to estimating the distribution of the T'^o statistic under the null hypothesis. Just as in the univariate case, this produces a single p-value:

$$p(T'^o) = \frac{1}{N} \sum_{k=1}^{N} H(T'^k, T'^o) .\tag{6}$$

It is this nonparametric approach that corrects for correlation among the tests, even without explicit diagonal entries in the covariance matrix.

We return to our example from the previous section. The U^k vectors are plotted in Fig. 3a, along with the $\alpha = 0.95$ decision boundary, and our sample is shown to lie outside of it. As can be seen, equal power is assigned to alternatives lying at the same distance from the origin in this space. Figure 3b shows this boundary mapped back into the space of the original p-values. The p-values of the individual partial tests are 0.36 and 0.022, and the combined result is 0.049.

2.4 Relation to Other Testing Procedures

The entire procedure is very similar to procedures used in correction for multiple tests, such as that proposed by Pantazis et al. [4]. In fact, another alternative
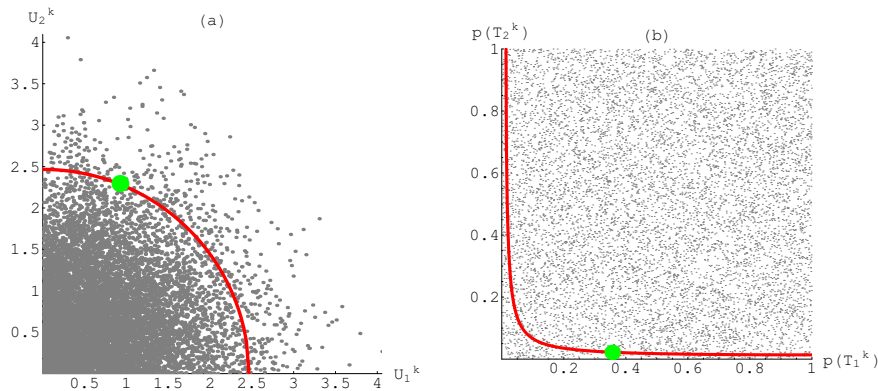
Fig. 3. The empirical distribution of our example plotted against the decision boundary at $\alpha = 0.95$. (a) The distribution of the U^k vectors, where the cutoff is a circle centered around the origin. (b) The distribution of the original p-values with the decision boundary pulled back into this space.

for a combining function is Tippet's $T'^k = \max_{j=1}^{M}(1 - p(T_j^k))$, which results in a Bonferroni-style correction [15]. Some authors have suggested methods of direct combination applied to the T_j^k statistics themselves [5, 6]. They are more appealing computationally, being $O(nMN)$ instead of our method's $O(nMN \log(N))$, but they do not avoid problems of differing scale or strong correlation.

Consider what happens when $T'^k = \sqrt{(T_1^k)^2 + (T_2^k)^2}$. Now, the first test dominates the results, and the overall p-value becomes 0.34. With $n_1 = n_2 = 100$ samples, our test becomes much more significant ($p = 0.0008$), while the direct combination test becomes even worse ($p = 0.44$).

3   Experimental Data and Results

The data for our experiments comes from a twin pair schizophrenia study conducted by Weinberger et al. [17]. High resolution ($0.9375 \times 0.9375 \times 1.5$ mm^3) Magnetic Resonance Imaging (MRI) scans were acquired from three different subject groups: 9 healthy monozygotic twin pairs (MZ), 10 healthy dizygotic twin pairs (DZ), and 9 monozygotic twin pairs with one twin discordant for schizophrenia and one twin unaffected. See Fig. 1 for some examples. A fourth group of 10 healthy non-related subject pairs (NR) was constructed by matching unrelated members of the two healthy groups. All four groups were matched for age, gender, and handedness. A tenth healthy, monozygotic twin pair was discarded due to segmentation problems attributed to head trauma suffered by one of the twins in a car accident at age seven. A tenth twin pair discordant for schizophrenia was discarded due to hydrocephaly in the unaffected twin.

The left and right lateral ventricles were segmented using supervised classification and 3-D connectivity [18]. An automatic morphological closing operation was applied to ensure a spherical topology. An area-preserving map was used to map them to a sphere, after which they were converted to a spherical harmonics representation (SPHARM) [19]. Correspondence on the boundary was established using the first order harmonics [20]. Point Distribution Models (PDMs) were constructed by uniformly sampling the boundary at corresponding points. The m-rep models were constructed using a robust method that ensures a common medial topology [21]. For our data, this consists of a single medial sheet with a 3×13 grid of medial atoms, which provides 98% volume overlap with the original segmentations.

From this data set, we wish to determine if the twin pairs that were more closely related had smaller variations in shape. We also wish to see if the shape variations between the discordant and the unaffected twins in the schizophrenic pairs is similar to the normal variation between healthy monozygotic twins. For this purpose, we use the partial test statistics:

$$T_j(x_1, y_1, x_2, y_2) = \frac{1}{n_2} \sum_{i=1}^{n_2} d(x_{2,i,j}, y_{2,i,j}) - \frac{1}{n_1} \sum_{i=1}^{n_1} d(x_{1,i,j}, y_{1,i,j}) \ . \qquad (7)$$

Here (x_1, y_1) form the twin pairs for one group, while (x_2, y_2) form the twin pairs for the other. The partial tests are applied separately to all three components of the medial atom location, $\mathbf{x}$, as well as the radius and two spoke directions. This gives six partial tests per medial atom, for a total of $M = 3 \times 13 \times 6 = 234$, much larger than the sample size. Each is a one-sided test that the variability in group 2 is larger than that in group 1.

For consistency with previous studies [22], all shapes were volume normalized. After normalization, we also applied m-rep alignment, as described by Fletcher et al. [2], to minimize the sum of squared geodesic distances between models in a medial analog of Procrustes alignment. First, the members of each twin pair were aligned with each other, and then the pairs were aligned together as a group, applying the same transformation to each member of a single pair.

In order to ensure invariance to rotations, we had to choose data-dependent coordinate axes for the $\mathbf{x}$ component of each medial atom. Our choice was the axes which diagonalized the sample covariance of the displacement vectors from one twin's atom position to the other at each site. While this had some influence on the results, the general trend was the same irrespective of the axes used.

For each pair of twin groups, we generated $N = 50,000$ permutations, and computed their p-value vectors using (4). Following Sect. 2.3, these were mapped into U^k vectors, from which the empirical distribution of the combined test statistic T'^k from (5) was estimated, producing a single global p-value via (6).

The results are summarized in Table 1. For comparison, we list the results of a previous study which used a univariate test on the average distance between corresponding points on the PDMs [22]. While we note that the significance of a p-value on an experimental data set is not a useful metric for comparing different methods, it is interesting to see the differences between the two. Our

tests give a consistent ranking: MZ ≈ DS < DZ ≈ NR, which is fully transitive. The boundary study, however, finds a significant difference between DZ and NR, but fails to identify the difference between DS and DZ.

	Our Study		Boundary Study [22]	
	Left	Right	Left	Right
MZ vs. DS	0.12	0.38	0.28	0.68
MZ vs. DZ	**0.00006**	**0.0033**	**0.0082**	**0.0399**
MZ vs. NR	**0.00002**	**0.00020**	**0.0018**	**0.0006**
DS vs. DZ	**0.020**	**0.0076**	0.25	0.24
DS vs. NR	**0.0031**	**0.00026**	**0.018**	**0.0026**
DZ vs. NR	0.16	0.055	**0.05**	**0.016**

Table 1. p-values for paired tests for the difference in the amount of shape variability in groups with different degrees of genetic similarity. Results from our method are in the first two columns, while results from a previous study [22] are in the last two for comparison. Groups are: monozygotic (MZ), monozygotic twins with one twin discordant for schizophrenia (DS), dizygotic (DZ), and non-related (NR). Results significant at the $\alpha = 0.95$ level are shown in **bold**.

We also performed local tests, to identify specific medial atoms with with strong differences. A multivariate test was conducted using our procedure on the 6 components of $\mathcal{M}(1)$ for each atom, and the results were corrected for multiple tests using the minimum p-value distribution across the shape, as described by Pantazis et al. [4]. The results are shown in Fig 4.

4   Conclusion

We have presented a true multivariate permutation test approach for hypothesis testing in direct products of metric spaces. The resulting test does not require a priori scaling factors to be chosen, and captures the true multivariate nature of the data. It is well-defined even in the high-dimensional, low-sample size case. The method has been applied to shape discrimination using m-reps, though it is suitable for any type of metric data, including potentially categorical data.

An important area for future research is the design of suitable partial tests to use in each space. Because they cannot be broken into smaller pieces than a single component of the direct product, the distance to the mean and similar tests are limited in the types of distributions they can describe. For example, the distance from the mean can only characterize an isotropic distribution on the sphere. An interesting candidate is the test designed by Hall and Tajvidi, which can test for equality of entire distributions in a single metric space [23]. This would allow us to relax our assumption of identical distribution about the mean. For manifolds, another possibility is the use of tests based on Distance
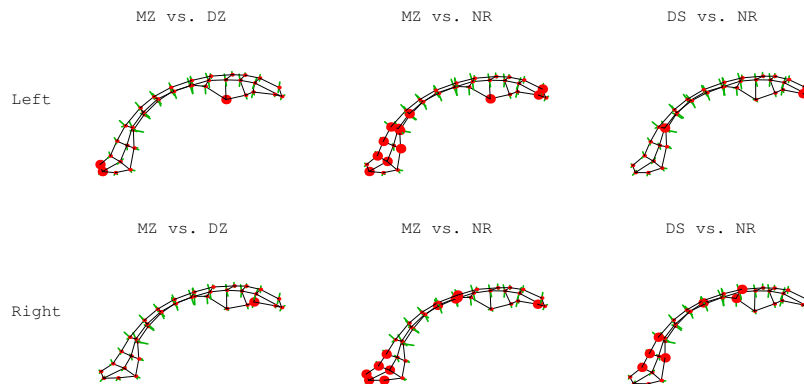
Fig. 4. Results for local tests for the difference in shape variability in groups with different degrees of genetic similarity. Atoms with differences significant at the $\alpha = 0.95$ level are shown in a larger size. Tests not shown had no significant local differences.

Weighted Discrimination [24]. It is also possible to extend this to different shape models, such as PDMs with surface normals or deformable templates [25].

Acknowledgments

References

1. Fletcher, P.T., Lu, C., Joshi, S.C.: Statistics of shape via principal component analysis on Lie groups. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR'03), Los Alamitos, CA (2003) 95–101
2. Fletcher, P.T., Lu, C., Pizer, S.M., Joshi, S.C.: Principal geodesic analysis for the study of nonlinear statistics of shape. IEEE Transactions on Medical Imaging **23** (2004) 995–1005
3. Fletcher, P.T.: Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI. PhD thesis, UNC at Chapel Hill (2004)
4. Pantazis, D., Leahy, R.M., Nichols, T.E., Styner, M.: Statistical surface-based morphometry using a non-parametric approach. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'04). (2004) 1283–1286
5. Chung, J.H., Fraser, D.A.S.: Randomization tests for a multivariate two-sample problem. Journal of the American Statistical Association **53** (1958) 729–735
6. Blair, R.C., Higgins, J.J., Karniski, W., Kromrey, J.D.: A study of multivariate permutation tests which may replace Hotelling's T^2 test in prescribed circumstances. Multivariate Behavioral Research **29** (1994) 141–164

7. Joshi, S.C., Fletcher, P.T., Yushkevich, P.A., Thall, A., Marron, J.S.: Multi-scale deformable model segmentation and statistical shape analysis using medial descriptions. IEEE Transactions on Medical Imaging **21** (2002) 538–550

8. Fréchet, M.: Les éléments aléatoires de nature quelconque dans un espace distancié. Annales de L'Institut Henri Poincaré **10** (1948) 215–310

9. Karcher, H.: Riemannian center of mass and mollifier smoothing. Communications on Pure and Applied Math **30** (1977) 509–541

10. Pennec, X.: Probabilities and statistics on Riemmanian manifolds: Basic tools for geometric measurements. In Cetin, A., Akarun, L., Ertuzun, A., Gurcan, M.N., Yardimci, Y., eds.: Proceedings of the IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99). (1999) 194–198

11. Kendall, W.S.: Probability, convexity, and harmonic maps with small image I: Uniqueness and fine existence. In: Proceedings of the London Mathematical Society. Volume 61. (1990) 371–406

12. Bradley, J.V.: Distribution-Free Statistical Tests. Prentice Hall, Englewood Cliffs, New Jersey (1968)

13. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping **15** (2002) 1–25

14. Anderson, T.W.: An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York (1958)

15. Pesarin, F.: Multivariate Permutation Tests with Applications in Biostatistics. John Wiley & Sons, Ltd., Chirchester (2001)

16. Pallini, A., Pesarin, F.: A class of combinations of dependent tests by a resampling procedure. In Jöckel, K.H., Rothe, G., Sendler, W., eds.: Bootstrapping and Related Techniques. Volume 376 of Lecture Notes in Economics and Mathematical Systems., Berlin, Springer-Verlag (1992) 93–97

17. Weinberger, D.R., Egan, M.F., Bertolino, A., Callicott, J.H., Mattay, V.S., Lipska, B.K., Berman, K.F., Goldberg, T.E.: Prefrontal neurons and the genetics of schizophrenia. Biological Psychiatry **50** (2001) 825–844

18. van Leemput, K., Maes, F., Vandermeulen, D., Seutens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Transactions on Medical Imaging **18** (1999) 897–908

19. Brechbühler, C., Gerig, G., Kübler, O.: Parameterization of closed surfaces for 3-D shape description. Computer Vision, Graphics, and Image Processing: Image Understanding **61** (1995) 195–170

20. Gerig, G., Styner, M., Jones, D., Weinberger, D.R., Lieberman, J.A.: Shape analysis of brain ventricles using SPHARM. In: Proceedings of the IEEE Workshop on Mathematical Methods in Biomedical Image Analysis. (2001) 171–178

21. Styner, M., Gerig, G., Joshi, S.C., Pizer, S.M.: Automatic and robust computation of 3-D medial models incorporating object variability. International Journal of Computer Vision **55** (2003) 107–122

22. Styner, M., Lieberman, J.A., McClure, R.K., Weinberger, D.R., Jones, D.W., Gerig, G.: Morphometric analysis of lateral ventricles in schizophrenia and healthy controls regarding genetic and disease-specific factors. Proceedings of the National Academy of Science **102** (2005) 4872–4877

23. Hall, P., Tajvidi, N.: Permutation tests for equality of distributions in high-dimensional settings. Biometrika **89** (2002) 359–374

24. Marron, J.S., Todd, M.J.: Distance weighted discrimination. Technical Report 1339, Operations Research and Industrial Engineering, Cornell University (2002)

25. Grenander, U., Keenan, D.M.: Towards automated image understanding. Advances in Applied Statistics: Statistics and Images **1** (1993) 89–103