Separability tests for high-dimensional, low sample size multivariate repeated measures data

Sean L. Simpson[a,b,*], **Lloyd J. Edwards**[b], **Martin A. Styner**[c], and **Keith E. Muller**[d]

[a]Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC 27157-1063

[b]Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420

[c]Departments of Psychiatry and Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7160

[d]Department of Health Outcomes and Policy, University of Florida, Gainesville, FL 32610-0177

Abstract

Longitudinal imaging studies have moved to the forefront of medical research due to their ability to characterize spatio-temporal features of biological structures across the lifespan. Valid inference in longitudinal imaging requires enough flexibility of the covariance model to allow reasonable fidelity to the true pattern. On the other hand, the existence of computable estimates demands a parsimonious parameterization of the covariance structure. Separable (Kronecker product) covariance models provide one such parameterization in which the spatial and temporal covariances are modeled separately. However, evaluating the validity of this parameterization in high-dimensions remains a challenge. Here we provide a scientifically informed approach to assessing the adequacy of separable (Kronecker product) covariance models when the number of observations is large relative to the number of independent sampling units (sample size). We address both the general case, in which unstructured matrices are considered for each covariance model, and the structured case, which assumes a particular structure for each model. For the structured case, we focus on the situation where the within subject correlation is believed to decrease exponentially in time and space as is common in longitudinal imaging studies. However, the provided framework equally applies to all covariance patterns used within the more general multivariate repeated measures context. Our approach provides useful guidance for high dimension, low sample size data that preclude using standard likelihood based tests. Longitudinal medical imaging data of caudate morphology in schizophrenia illustrates the approaches appeal.

Keywords

Kronecker product; Separable Covariance; Multivariate repeated measures; Likelihood ratio test; Spatio-temporal data; Linear exponent autoregressive model

---

[*]Correspondence to: Sean L. Simpson, Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, NC 27157-1063, slsimpso@wakehealth.edu.

1. Introduction

Multivariate repeated measures studies are characterized by data that have more than one set of correlated outcomes or repeated factors. Spatio-temporal data fall into this more general category since the outcome variables repeat in both space and time. Valid analysis requires accurately modeling the correlation pattern. Muller et al. (2007) and Gurka et al. (2011) showed that under-specifying the correlation structure can severely inflate test size in inference about fixed effects in the general linear mixed model. With multivariate repeated measures data, modeling the correlation pattern separately for each repeated factor has substantial advantages. Most importantly, the approach allows choosing and tuning each model separately which improves accuracy and makes model fitting easier. Furthermore, the approach inherently allows using fewer parameters than does an unstructured model. Use of the Kronecker product provides an appealing way to combine these factor-specific correlation structures into an overall correlation model. No additional parameters are needed to combine any mathematically valid correlation patterns into a valid overall pattern.

Galecki (1994) gave a detailed treatment of Kronecker product covariance structures, also known as separable covariance models. A covariance matrix is *separable* if and only if it can be written as $\Sigma = \Gamma \otimes \Omega$, where Γ and Ω are factor specific covariance matrices (e.g. the covariance matrices for the temporal and spatial dimensions of spatio-temporal data respectively). A key advantage of the model lies in the ease of interpretation in terms of the independent contribution of every repeated factor to the overall within-subject error covariance matrix. The model also accommodates correlation matrices with nested parameter spaces and factor specific within-subject variance heterogeneity. Galecki (1994), Naik and Rao (2001), and Mitchell et al. (2006) detailed the computational advantages of the Kronecker product covariance structure. The partial derivatives, inverse, and Cholesky decomposition of the overall covariance matrix can be performed more easily on the smaller dimensional factor specific models.

Limitations of separable models have been noted by various authors. Most importantly, as mentioned by Cressie and Huang (1999), patterns of interaction among the various factors cannot be modeled when utilizing a Kronecker product structure. Galecki (1994), Huizenga et al. (2002), and Mitchell et al. (2006) all noted that a lack of identifiability can result with such a model. The indeterminacy stems from the fact that if $\Sigma = \Gamma \otimes \Omega$ is the overall within-subject error covariance matrix, Γ and Ω are not unique since for $a \neq 0$, $a\Gamma \otimes (1/a)\Omega = \Gamma \otimes \Omega$. However, this nonidentifiability can be fixed by rescaling one of the factor specific covariance matrices so that one of its diagonal nonzero elements is equal to 1. With homogeneous variances, the rescaled matrix is a correlation matrix. It is also important to note that within a given subject all factors must have *consistently-spaced* measurements. In the context of spatio-temporal data this means that at each time point a given subject must have the same number of measurements taken at the same spatial locations.

Several tests have been developed to determine the validity of assuming a separable covariance model. General (pure) tests use unstructured null and alternative hypothesis matrices. Shitan and Brockwell (1995) constructed an asymptotic chi-square test for general separability. Likelihood ratio tests for general separability were derived by Lu and

Zimmerman (2005), Mitchell et al. (2006), and Roy and Khattree (2003). Fuentes (2006) developed a general test for separability of a spatio-temporal process utilizing spectral methods.

Structure-specific tests of separability have particular structure assumed for the null hypothesis but generally not for the alternative hypothesis. Structured tests of separability have been proposed by Roy and Khattree (2005a, 2005b) and Roy and Leiva (2008). Roy and Khattree (2005a) derived a test for the case with one factor matrix being compound symmetric and the other unstructured. Roy and Khattree (2005b) developed a test for when one factor specific matrix has the discrete-time AR(1) structure and the other is unstructured. The test of Roy and Leiva (2008) requires either a compound symmetric or discrete-time AR(1) structure for the factor specific matrices. Simpson (2010) developed an adjusted likelihood ratio test of two-factor separability for unbalanced multivariate repeated measures data. The approach can be generalized to factor specific matrices of any structure.

All of the authors just mentioned noted that none of the separability tests developed thus far can handle high-dimensional, low sample size (HDLSS) data due to the potential nonexistence of a computable estimate for an unstructured covariance fit (the alternative hypothesis). We provide a scientifically informed approach to conducting useful tests of separability in the presence of HDLSS, a common problem in medical imaging and various kinds of "-omics" data. We illustrate our method with data concerning caudate morphology in schizophrenia. The data come from longitudinal MRI scans of the left caudate for 240 schizophrenia patients and 56 controls. The surface of each object was parameterized via the m-rep method as described in Styner and Gerig (2001). The caudate shape was determined as a 3×7 grid of mesh points (see Figure 1). Data were reduced to one outcome measure: *radius* in cm as a measure of local object width (21 locations per caudate with no missing data). The distance between two radii for a given subject was calculated as the mean Euclidian distance over all images. Scans were taken up to 47 months post-baseline with the median and maximum number of scans per subject being 3 and 7 respectively.

We present our approach to testing for separability in HDLSS data in section 2. Simulation studies in section 3 help evaluate the approach. In section 4 the caudate morphology data illustrate the use of the proposed testing procedure. We conclude with a summary discussion including planned future research in section 5.

2. Likelihood ratio tests of separability

We examine both a structured and general test of separability to illustrate our approach. We consider the following structured likelihood ratio test of separability for the Kronecker product *linear exponent autoregressive* (KP LEAR) model which has been shown to work well for situations in which the within subject correlation is believed to decrease exponentially in time and space (Simpson et al., 2010, 2013). Suppose y_i is a $t_i s_i \times 1$ vector of $t_i s_i$ observations (e.g., t_i temporal measurements and s_i spatial measurements) on the ith subject $i \in \{1, \ldots, N\}$. Here $c(y_{ijl}, y_{ikl}) = \rho_{i\gamma;jk}$ and $c(y_{ijl}, y_{ijm}) = \rho_{i\omega;lm}$ represent the temporal (or factor 1) and spatial (or factor 2) correlations respectively, for $c(\cdot)$ the correlation operator. Then for $\Gamma_i = \{\rho_{i\gamma;jk}\}$ (the temporal/factor 1 correlation matrix) and Ω_i

= $\{\rho_{i\omega;lm}\}$ (the spatial/factor 2 correlation matrix), the factor specific *linear exponent autoregressive* (LEAR) correlation structures are

$$\rho_{i\gamma;jk}=\mathscr{C}(y_{ijl},y_{ikl})=\begin{cases} \rho_\gamma^{d_{t;\,\min}+\delta_\gamma[(d(t_{ijl},t_{ikl})-d_{t;\,\min})/(d_{t;\,\max}-d_{t;\,\min})]} & j\neq k \\ 1 & j=k \end{cases},\quad (1)$$

$$\rho_{i\omega;lm}=\mathscr{C}(y_{ijl},y_{ijm})=\begin{cases} \rho_\omega^{d_{s;\,\min}+\delta_\omega[(d(s_{ijl},s_{ijm})-d_{s;\,\min})/(d_{s;\,\max}-d_{s;\,\min})]} & l\neq m \\ 1 & l=m \end{cases}.\quad (2)$$

where $d(t_{ijl}, t_{ikl})$ and $d(s_{ijl}, s_{ikl})$ are the distances between measurement times and locations respectively. In turn $(d_{t;\min}, d_{s;\min})$ and $(d_{t;\max}, d_{s;\max})$ are computational *constants* equal to the minimum and maximum number of temporal and spatial distance units across all subjects. Parameters ρ_γ and ρ_ω are the correlations between observations separated by one unit of time and distance respectively, and δ_γ and δ_ω are the decay speeds. We assume 0 ρ_γ, $\rho_\omega < 1$ and 0 δ_γ, δ_ω. The $(d_{t;\min}, d_{s;\min})$ and $(d_{t;\max}, d_{s;\max})$ constants allow the model to adapt to the data and scale distance such that the multiplier of the decay speeds δ_γ and δ_ω, $(d(t_{ijl}, t_{ikl}) - d_{t;\min})/(d_{t;\max} - d_{t;\min})$ and $(d(s_{ijl}, s_{ijm}) - d_{s;\min})/(d_{s;\max} - d_{s;\min})$, is between 0 and 1 for computational purposes. One could also consider tuning the constants if necessary to address, for example, convergence issues. Simpson et al. (2010, 2013) contain further details of the model. Note that each factor specific LEAR model can also be reparameterized as an exponential model (or continuous-time AR(1) model) with a multiplicative nugget effect (see Schabenberger and Gotway, 2005 for details on nugget effects).

Following the preceding notation, and assuming that $y_i|X_i \sim N_{t_i s_i}(\mu_i = X_i\beta, \Sigma_i = \sigma^2[\Gamma_i(\tau_\gamma) \otimes \Omega_i(\tau_\omega)])$ and is independent of $y_i|X_i$ for i i' where Γ_i and Ω_i are defined in equations 1 and 2 with $\tau = \{\tau_\gamma; \tau_\omega\} = \{\delta_\gamma, \rho_\gamma; \delta_\omega, \rho_\omega\}$, the structured likelihood ratio test of separability for the KP LEAR model is

$$H_0: \textstyle\sum_i = \sigma^2\Gamma_i \otimes \Omega_i; \Gamma_i, \Omega_i \text{ LEAR}$$
$$\text{vs.} \qquad\qquad (3)$$
$$H_1: \textstyle\sum_i \text{ unstructured, positive definite (PD)}.$$

The log-likelihood function of the parameters given the data under H_0 is

$$\begin{aligned} l(y; \beta, \sigma^2, \tau) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln|\sigma^2\Gamma_i \otimes \Omega_i| - \frac{1}{2\sigma^2}\sum_{i=1}^{N}r_i(\beta)'(\Gamma_i \otimes \Omega_i)^{-1}r_i(\beta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\left(t_i s_i \ln(\sigma^2) + \ln|\Gamma_i \otimes \Omega_i|\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}r_i(\beta)'(\Gamma_i \otimes \Omega_i)^{-1}r_i(\beta), \end{aligned} \qquad (4)$$

where $n = \sum_{i=1}^{N}t_i s_i$ and $r_i(\beta) = y_i - X_i\beta$. The log-likelihood function of the parameters given the data under H_1 is

$$l(\boldsymbol{y};\boldsymbol{\beta},\sigma^2,\boldsymbol{\tau})=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}^{N}\ln\left|\sum_i\right|-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{r}_i(\boldsymbol{\beta})'\sum_i^{-1}\boldsymbol{r}_i(\boldsymbol{\beta}) \quad (5)$$

The standard likelihood ratio is given by

$$\Lambda=\frac{\max_{H_0}L}{\max_{H_1}L}$$

$$=\frac{\exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_0)'(\hat{\boldsymbol{\Gamma}}_i^{-1}\otimes\hat{\boldsymbol{\Omega}}_i^{-1})\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_0)\right\}\prod_{i=1}^{N}(2\pi\hat{\sigma}^2)^{-t_i s_i/2}|\hat{\boldsymbol{\Gamma}}_i|^{-s_i/2}|\hat{\boldsymbol{\Omega}}_i|^{-t_i/2}}{\exp\left\{-\frac{1}{2}\sum_{i=1}^{N}\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_1)'\hat{\sum}_i^{-1}\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_1)\right\}\prod_{i=1}^{N}(2\pi)^{-t_i s_i/2}|\hat{\sum}_i|^{-1/2}}$$

$$=\frac{\exp\left\{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{N}\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_0)'(\hat{\boldsymbol{\Gamma}}_i^{-1}\otimes\hat{\boldsymbol{\Omega}}_i^{-1})\boldsymbol{r}_i(\hat{\boldsymbol{\beta}}_0)\right\}\hat{\sigma}^{-n}\prod_{i=1}^{N}|\hat{\boldsymbol{\Gamma}}_i|^{-s_i/2}|\hat{\boldsymbol{\Omega}}_i|^{-t_i/2}}{\exp\{-n/2\}\prod_{i=1}^{N}|\hat{\sum}_i|^{-1/2}}, \quad (6)$$

where L is the likelihood function and the maximum likelihood (ML) estimates are derived following the approach in Simpson (2010) and Simpson et al. (2013). Namely, given the imbalance in the data, the Newton-Raphson algorithm is employed to simultaneously solve for the estimates of Σ_i and β. More specifically, for the unstructured case, each estimate of Σ_i is a subset of the estimate for the overall general covariance matrix Σ in which Σ_i is embedded. For example, the caudate morphology data discussed in the Introduction has 7 total time points and 21 spatial locations leading to an overall Σ to be estimated of dimensions 147×147. If the first subject is measured at times 1, 3, and 7, then its covariance matrix Σ_1 consists of the 1st, 3rd, and 7th 21×21 main diagonal blocks of Σ. Under regularity conditions, $-2\ln\Lambda$ is asymptotically distributed as a χ_ν^2 random variable. The associated degrees of freedom parameter ν is given by

$$\nu=\left(\frac{t_{\text{tot}}s_{\text{tot}}(t_{\text{tot}}s_{\text{tot}}+1)}{2}\right)-5, \quad (7)$$

where t_{tot} and s_{tot} equal the total number of time points and spatial locations respectively. Here we use the adjusted LRT (aLRT) of Simpson (2010), namely $-2\ln\Lambda \approx k\chi_\nu^2$, where

$$k=N/(N-t_{\text{tot}}s_{\text{tot}}) \quad (8)$$

and conduct several tests of "marginal KP LEARness" since the number of observations per subject precludes conducting an overall aLRT using all of the data.

We also consider the following general likelihood ratio test of separability:

$$H_0 : \textstyle\sum_i = \boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_i ; \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i \text{ unstructured}$$
$$\text{vs.} \tag{9}$$
$$H_1 : \textstyle\sum_i \text{ unstructured, PD}$$

We again employ the aLRT previously defined in equations 6 and 8, but whose formula now reduces to

$$\Lambda = \frac{\max_{H_0} L}{\max_{H_1} L} = \frac{\hat{\sigma}^{-n} \prod\limits_{i=1}^{N} |\hat{\boldsymbol{\Gamma}}_i|^{-s_i/2} |\hat{\boldsymbol{\Omega}}_i|^{-t_i/2}}{\prod\limits_{i=1}^{N} |\hat{\textstyle\sum}_i|^{-1/2}}, \tag{10}$$

given that the matrices $\boldsymbol{\Gamma}_i$ and $\boldsymbol{\Omega}_i$ are unstructured. The associated degrees of freedom ν is now given by

$$\nu = \left(\frac{t_{\text{tot}} s_{\text{tot}} (t_{\text{tot}} s_{\text{tot}} + 1)}{2} \right) - \left(\frac{t_{\text{tot}} (t_{\text{tot}} + 1)}{2} + \frac{s_{\text{tot}} (s_{\text{tot}} + 1)}{2} - 1 \right). \tag{11}$$

As stated previously, we would like to provide a scientifically informed approach to assessing the appropriateness of a separable model with high-dimensional data. To do this we will conduct c marginal aLRT tests using subsets of the data corresponding to diagonal blocks of the covariance matrices. The approach covers a large fraction of the covariance space and is especially useful in situations with most of the information contained along diagonal blocks (i.e., correlation dies out along off-diagonals), which is the case for our longitudinal imaging example concerning caudate morphology in schizophrenia. For balanced data, each diagonal subset will contain data from all subjects. However, for unbalanced data, subsets should either be chosen so that the same number of subjects are used in each test or some weighting of the c tests should be considered based on the number of subjects in each. We take the former approach in the simulation studies and data application that follow. After the c aLRT tests are conducted, a false discovery rate correction is applied that controls for multiple testing given dependent tests (Benjamini and Yekateuli, 2001) and significance of the overall test is declared if any of the c p-values is significant. The dependencies among covariance parameters in samples from a multivariate Gaussian population were given by Wishart (1928).

This marginal testing approach can be thought of as a generalization of the more formalized framework of Molenberghs et al. (2011). They presented a pseudo-likelihood based method to partition prohibitively large data sets into M sub-samples, analyze each partition member, and combine the results across partitions for parameter estimation. More formally, the full sample is broken into M sub-samples of size n_m, where $m \in \{1, \ldots, M\}$. The pseudo-log-likelihood for sample m is

$$pl_m(\boldsymbol{\theta}_m) = \sum_{i=1}^{n_m} l(\boldsymbol{y}_{mi}|\boldsymbol{\theta}_m), \quad (12)$$

where $l(\cdot)$ is the log-likelihood that would be considered if the mth sub-sample were the entire data set. In our case, $l(\cdot) = l(\mathbf{y}_i; \boldsymbol{\beta}, \sigma^2, \boldsymbol{\tau})$ from equation 4 for the sub-sample under H_0 for the structured test and $l(\cdot) = l(\mathbf{y}_i; \boldsymbol{\beta}, \boldsymbol{\Gamma}_i, \boldsymbol{\Omega}_i)$ under H_0 for the general test. Under H_1, $l(\cdot) = l(\mathbf{y}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma}_i)$ from equation 5 for both the structured and general test. Extending their approach directly into our context would involve averaging test statistics or p-values across sub-samples (as they do for estimation of $\boldsymbol{\theta}$) and drawing inference from the averaged value. Our approach is slightly different in that we test for marginal separability for each of the M sub-samples and draw inference based on the number of these M tests that are significant. Our method provides robustness against outlying sub-sample p-values.

3. Simulation studies

3.1 Structured tests

To assess the empirical performance of the structured aLRT in equation 3, we conducted two simulation studies. The objective of the first study was to assess how much information is lost in taking the diagonal subset approach. We simulated multivariate repeated measures data with a 16×16 covariance matrix under H_0 (two 4×4 factor specific matrices) and H_1 and then conducted 1) an aLRT using all of the data, and 2) four tests using diagonal 2×2 subsets of the factor specific matrices following the testing procedure delineated in the previous section (subset aLRT). The data were generated under H_0 with $\boldsymbol{\rho} = [\rho_\gamma \, \rho_\omega]' = [0.8 \; 0.8]'$, $\boldsymbol{\delta} = [\delta_\gamma \, \delta_\omega]' = [(d_{t;\max} - d_{t;\min})/4 \; (d_{s;\max} - d_{s;\min})/4]'$, $\sigma^2 = 1$, and two-unit distance intervals for both factors (space and time). The data were generated under H_1 based on Theorem 10.13 of Muller and Stewart (2006). Simulated test size and power at the $\alpha = 0.05$ level was examined for tests 1) and 2) with sample sizes of $N = 40, 80, 120, 160,$ and 200. Without loss of generality, the mean model was set to $\beta = 0$ (one group with mean 0) (Simpson, 2010). Each simulation consisted of 5,000 realizations.

Table 1 shows the results of this first simulation study for the structured tests. It contains the simulated test sizes and power for the full aLRT and subset aLRT approaches. For the subset approach, the table also shows the test size and power by the number (out of $c = 4$) of significant p-values required for overall significance to be declared. There is very slight test size inflation for the subset aLRT when one significant p-value is required. Test size is well controlled when more than one significant p-value is required as it is for the full data aLRT. Additionally, there is no loss in power for the subset aLRT as compared with the full data aLRT. Although the subset method is for situations in which a full data test is infeasible, this comparison gives us confidence that little information is lost when taking our subset approach.

The objective of the second simulation study was to assess the type I error rate and power of the subset aLRT approach in a higher-dimensional setting in which current separability tests are unsuitable due to the potential nonexistence of a computable estimate for an unstructured covariance fit (while theoretical existence just requires $N > t_{\text{tot}} s_{\text{tot}}$, $N \gg t_{\text{tot}} s_{\text{tot}}$ is generally

required for computational existence due to estimation algorithm nonconvergence in relatively smaller sample size settings). Our simulations were aimed to mimic the schizophrenia and caudate morphology data discussed in Simpson et al. (2013) and in the next section. To do this, we again generated data under H_0 with $\boldsymbol{\rho} = [\rho_\gamma \, \rho_\omega]' = [0.8 \; 0.8]'$, $\boldsymbol{\delta} = [\delta_\gamma \, \delta_\omega]' = [(d_{t;\max} - d_{t;\min})/4 \; (d_{s;\max} - d_{s;\min})/4]'$, and $\sigma^2 = 1$, and under H_1 based on Theorem 10.13 of Muller and Stewart (2006). To mimic the example, the data generated were unbalanced with $\max_i(s_i) = s_{\text{tot}} = s = 21$, $\max_i(t_i) = t_{\text{tot}} = 7$, and $\text{med}_i(t_i) = 3$. There were $(t_i \cdot s) \, \varepsilon \, [21, 147]$ observations per subject ($\frac{1}{10}$ had 21 observations, $\frac{2}{5} \, (+2)$ had 63 observations, $\frac{2}{5} \, (-2)$ had 105 observations, and $\frac{1}{10}$ had 147 observations), each at two-unit distance intervals. Ten subset tests of 2×2 diagonal blocks of the spatial matrix ($\boldsymbol{\Omega}$) were conducted using the entire 7×7 temporal matrix $\boldsymbol{\Gamma}_i$ (thus, for each subset test, $\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}_{\text{subset}}$ was 14×14). Ten tests occur since there are ten 2×2 diagonal blocks in the 21×21 spatial matrix ($\boldsymbol{\Omega}$) with 1 diagonal element (the last element) omitted. Given the imbalance in temporal measurements, the approach ensures that the same number of subjects are used in each subtest. Simulated test size and power at the $\alpha = 0.05$ level was examined for sample sizes of $N = 140, 240, 280,$ and 320. The sample size of $N = 140 < t_{\text{tot}} s_{\text{tot}} = 147$ was chosen to assess the performance of the test in the context of a theoretically nonexistent unstructured covariance estimate. The other sample sizes were chosen to mimic the schizophrenia and caudate morphology data.

The results of the second simulation study for the structured tests are shown in Table 2. It contains the simulated test size and power for the subset aLRT by the number (out of $c = 10$) of significant p-values required for overall significance to be declared. There is severe test size inflation for all sample sizes when only 1 significant p-value is required. However, the test size becomes controlled for $N = 280$ and 320 when 2 or more significant values are required, for $N = 240$ when 3 or more are required, and for $N = 140$ when 9 or 10 is required. Regardless of the number of significant subtest p-values required for overall significance, the test remains extremely powerful.

3.2 General tests

To assess the empirical performance of the unstructured aLRT in equation 9, we conducted the same two simulation studies detailed in the previous subsection for the structured case. Here the data were generated under H_0 and H_1 based on Theorem 10.13 of Muller and Stewart (2006). Table 1 exhibits the results of the first simulation study for the general tests assessing information loss when taking the diagonal subset approach. Test size is well controlled for the subset aLRT across all conditions as it is for the full data aLRT. Moreover, there is minimal to no loss in power for all sample sizes with the subset aLRT when one significant p-value is required. However, the power loss increases as the number of significant p-values required increases, with this effect mitigated at larger sample sizes.

Table 2 displays the results of the second simulation study for the general tests assessing the type I error rate and power of the subset aLRT approach in the higher-dimensional setting. Severe test size inflation occurs for all sample sizes when only 1 significant subtest p-value is required and remains above the $\alpha = 0.05$ level until 6 or more (out of $c = 10$) significant values are required for $N = 240, 280,$ and 320, and until 8 or more are required for $N = 140$.

As with the structured test, the unstructured subset aLRT remains extremely powerful for all parameter combinations.

4. Test of separability for schizophrenia and caudate morphology example

We model the schizophrenia and caudate morphology data discussed in the Introduction with an intercept-only general linear model for repeated measures (Simpson et al., 2013).

That is, $\boldsymbol{y}_i \sim N_{t_i s}(\boldsymbol{\mu}_i = \mathbf{1}'_{t_i s}\boldsymbol{\beta}_0, \sum_i)$, where the $\log_2$(radius) values for each of the $s_i = s = 21$ locations (spatial factor) and t_i images (temporal factor) for the ith subject ($i \in \{1, \ldots, 296\}$) are contained in $\boldsymbol{y}_i$ ($t_i \cdot 21 \times 1$). For the assessment of covariance structure separability, the 240 cases and 56 controls remain combined as the sample size of the controls precludes fitting an unstructured model (since $N_{\text{controls}} = 56 < t_{\text{tot}}s_{\text{tot}} = 147$) and thus precludes a standard test for homogeneity of covariance (e.g., Box's test). While a modified homogeneity test employing a subsampling approach like the one discussed here may prove useful in this context, an examination of this is beyond the scope of this work. This data example is meant to be illustrative of the benefits of our separability testing procedure.

As evidenced by information criteria and observed vs. predicted correlation plots, Simpson et al. (2013) showed that the Kronecker product LEAR model appears to provide a good fit to the caudate morphology data. However, the validity of the separable assumption should be assessed as there may be space × time interactions which cannot be modeled with the Kronecker structure. In order to test separability, Simpson (2010) was forced to reduce the data by picking four (out of the 21) representative spatial locations to accommodate the dimensions of the data. Here we apply our structured subset aLRT approach detailed in section 2 by conducting 10 subset tests utilizing the entire 7 × 7 temporal matrix $\boldsymbol{\Gamma}_i \otimes \boldsymbol{\Omega}$ 2 × 2 diagonal blocks of the spatial matrix as in the second simulation study. For a significance level of $\alpha = 0.05$ and, based on the simulations conducted in section 3.1, declare significance if 3 or more (out of $c = 10$) of the subset p-values are significant. The null hypothesis is rejected since more than 3 of the tests had significant p-values, which implies that the assumption of separability appears invalid in this case. In order to gain insight into this finding we examine the following estimates (each multiplied by 100) for 4 (of the 21) spatial locations and the 3 time points for subject $i = 4$ ($t_4 = 3$, $s = 4$) (also in Simpson, 2010):

$$\hat{\sigma}^2\hat{\Gamma}_4 \otimes \hat{\Omega} = \begin{pmatrix}
1.36 & 0.49 & 0.42 & 0.24 & 1.08 & 0.39 & 0.34 & 0.19 & 1.07 & 0.38 & 0.33 & 0.19 \\
0.49 & 1.36 & 0.28 & 0.33 & 0.39 & 1.08 & 0.22 & 0.26 & 0.38 & 1.07 & 0.22 & 0.26 \\
0.42 & 0.28 & 1.36 & 0.18 & 0.34 & 0.22 & 1.08 & 0.14 & 0.33 & 0.22 & 1.07 & 0.14 \\
0.24 & 0.33 & 0.18 & 1.36 & 0.19 & 0.26 & 0.14 & 1.08 & 0.19 & 0.26 & 0.14 & 1.07 \\
1.08 & 0.39 & 0.34 & 0.19 & 1.36 & 0.49 & 0.42 & 0.24 & 1.08 & 0.39 & 0.34 & 0.19 \\
0.39 & 1.08 & 0.22 & 0.26 & 0.49 & 1.36 & 0.28 & 0.33 & 0.39 & 1.08 & 0.22 & 0.26 \\
0.34 & 0.22 & 1.08 & 0.14 & 0.42 & 0.28 & 1.36 & 0.18 & 0.34 & 0.22 & 1.08 & 0.14 \\
0.19 & 0.26 & 0.14 & 1.08 & 0.24 & 0.33 & 0.18 & 1.36 & 0.19 & 0.26 & 0.14 & 1.08 \\
1.07 & 0.38 & 0.33 & 0.19 & 1.08 & 0.39 & 0.34 & 0.19 & 1.36 & 0.49 & 0.42 & 0.24 \\
0.38 & 1.07 & 0.22 & 0.26 & 0.39 & 1.08 & 0.22 & 0.26 & 0.49 & 1.36 & 0.28 & 0.33 \\
0.33 & 0.22 & 1.07 & 0.14 & 0.34 & 0.22 & 1.08 & 0.14 & 0.42 & 0.28 & 1.36 & 0.18 \\
0.19 & 0.26 & 0.14 & 1.07 & 0.19 & 0.26 & 0.14 & 1.08 & 0.24 & 0.33 & 0.18 & 1.36
\end{pmatrix}$$

$$\hat{\Sigma}_4 = \begin{pmatrix}
1.40 & 0.54 & 0.52 & 0.31 & 1.09 & 0.44 & 0.55 & 0.19 & 1.05 & 0.35 & 0.65 & 0.23 \\
0.54 & 1.37 & 0.61 & 0.70 & 0.27 & 0.98 & 0.56 & 0.48 & 0.37 & 1.11 & 0.60 & 0.47 \\
0.52 & 0.61 & 2.00 & 0.34 & 0.65 & 0.55 & 1.80 & 0.21 & 0.77 & 0.61 & 1.88 & 0.27 \\
0.31 & 0.70 & 0.34 & 1.04 & 0.24 & 0.63 & 0.23 & 0.63 & 0.21 & 0.64 & 0.30 & 0.70 \\
1.09 & 0.27 & 0.65 & 0.24 & 1.63 & 0.52 & 0.75 & 0.21 & 1.22 & 0.28 & 0.81 & 0.19 \\
0.44 & 0.98 & 0.55 & 0.63 & 0.52 & 1.19 & 0.56 & 0.56 & 0.42 & 1.01 & 0.60 & 0.52 \\
0.55 & 0.56 & 1.80 & 0.23 & 0.75 & 0.56 & 2.14 & 0.16 & 0.82 & 0.61 & 1.93 & 0.20 \\
0.19 & 0.48 & 0.21 & 0.63 & 0.21 & 0.56 & 0.16 & 0.75 & 0.16 & 0.50 & 0.18 & 0.61 \\
1.05 & 0.37 & 0.77 & 0.21 & 1.22 & 0.42 & 0.82 & 0.16 & 1.48 & 0.47 & 0.91 & 0.22 \\
0.35 & 1.11 & 0.61 & 0.64 & 0.28 & 1.01 & 0.61 & 0.50 & 0.47 & 1.30 & 0.61 & 0.53 \\
0.65 & 0.60 & 1.88 & 0.30 & 0.81 & 0.60 & 1.93 & 0.18 & 0.91 & 0.61 & 2.19 & 0.24 \\
0.23 & 0.47 & 0.27 & 0.70 & 0.19 & 0.52 & 0.20 & 0.61 & 0.22 & 0.53 & 0.24 & 0.90
\end{pmatrix}.$$

The estimates show that a space × time interaction exists as the spatial covariance pattern (among the four caudate radii) changes across the three time points (the 3 4 × 4 blocks along the diagonal of $\hat{\Sigma_4}$). To further investigate whether the significant test results are due more to this space × time interaction or to heterogeneity of variance (our model assumes homogeneity) we reconduct the structured subset aLRT. Under H_0 we fit an unstructured variance model along with a LEAR correlation model for the spatial covariance matrix, Ω, and a single variance parameter (the observed temporal variances are constant) with a LEAR correlation model for the temporal covariance matrix, Γ_i. Again the null hypothesis is rejected implying that the space × time interaction is the main cause of the invalidity of the separability assumption. The separable LEAR model does provide a reasonable approximation to the completely unstructured model with 78 − 5 = 73 fewer parameters.

And with $\max_i(t_i s_i) = t_{tot} s_{tot} = 147$, the separable LEAR model has 10878 − 5 = 10873 fewer parameters than a completely unstructured model (for which convergence is currently impossible). However, given that our separability test results rendered the separability assumption invalid and the evident space × time interaction, a nonseparable model like that of Kim and Zimmerman (2012) or Fonseca and Steel (2011) should likely be employed.

5. Discussion

The subset aLRT approach provides a statistically reasonable approach to test the validity of the separability assumption when the standard tests do not apply. More specifically, the diagonal approach we take parsimoniously covers a large fraction of the covariance space

and is especially useful in situations like ours where most of the information is contained along diagonal blocks (i.e., correlation dies out along off-diagonals). As evidenced by the simulation results, the subset aLRT is a powerful test that also controls test size with careful selection of the number of significant subtests needed for overall significance.

Future work examining other covariance subspace sampling techniques will prove useful given the contextual nature of the problem. For example, conducting subtests based on random (as opposed to diagonal) subsets of the covariance matrix affords a method amenable to all covariance patterns regardless of whether most of the information is contained along the diagonal blocks. However, the approach may be more computationally intensive and less efficient when most of the information is in the middle of the matrix.

Acknowledgments

References

Benjamini Y, Yekateuli D. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics. 2001; 29:1165–1188.

Cressie N, Huang H. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association. 1999; 94:1330–1340.

Fonseca TCO, Steel MFJ. A General Class of Nonseparable Space-Time Covariance Models. Environmetrics. 2011; 22:224–242.

Fuentes M. Testing for separability of spatial-temporal covariance functions. Journal of Statistical Planning and Inference. 2006; 136:447–466.

Gurka MJ, Edwards LJ, Muller KE. Avoiding bias in mixed model inference for fixed effects. Statistics in Medicine. 2011; 30:2696–2707. [PubMed: 21751227]

Huizenga HM, de Munck JC, Waldorp LJ, Grasman RP. Spatiotemporal EEG/MEG source analysis based on a parametric noise covariance model. IEEE Transactions on Biomedical Engineering. 2002; 49:533–539. [PubMed: 12046698]

Kim C, Zimmerman DL. Unconstrained models for the covariance structure of multivariate longitudinal data. Journal of Multivariate Analysis. 2012; 107:104–118.

Lu N, Zimmerman DL. The likelihood ratio test for a separable covariance matrix. Statistics and Probability Letters. 2005; 73:449–457.

Mitchell MW, Genton MG, Gumpertz ML. A likelihood ratio test for separability of covariances. Journal of Multivariate Analysis. 2006; 97:1025–1043.

Molenberghs G, Verbeke G, Iddi S. Pseudo-likelihood methodology for partitioned large and complex samples. Statistics and Probability Letters. 2011; 81:892–901.

Muller KE, Edwards LJ, Simpson SL, Taylor DJ. Statistical tests with accurate size and power for balanced linear mixed models. Statistics in Medicine. 2007; 26:3639–3660. [PubMed: 17394132]

Muller, KE.; Stewart, PW. Linear Model Theory: Univariate, Multivariate, and Mixed Models. John Wiley & Sons, Inc; New Jersey: 2006.

Naik DN, Rao SS. Analysis of multivariate repeated measures data with a kronecker product structured covariance matrix. Journal of Applied Statistics. 2001; 28:91–105.

Roy A, Khattree R. Tests for mean and covariance structures relevant in repeated measures based discriminant analysis. Journal of Applied Statistical Science. 2003; 12:91–104.

Roy A, Khattree R. On implementation of a test for kronecker product covariance structure for multivariate repeated measures data. Statistical Methodology. 2005a; 2:297–306.

Roy A, Khattree R. Testing the hypothesis of a kronecker product covariance matrix in multivariate repeated measures data. Proceedings of the 30th Annual SAS Users Group International Conference (SUGI). 2005b

Roy A, Leiva R. Likelihood ratio tests for triply multivariate data with structured correlation on spatial repeated measurements. Statistics and Probability Letters. 2008; 78:1971–1980.

Schabenberger, O.; Gotway, CA. Statistical Methods for Spatial Data Analysis. Chapman & Hall; Boca Raton, FL: 2005.

Shitan M, Brockwell PJ. An asymptotic test for separability of a spatial autoregressive model. Communications In Statistics-Theory and Methods. 1995; 24:2027–2040.

Simpson SL. An adjusted likelihood ratio test for separability in unbalanced multivariate repeated measures data. Statistical Methodology. 2010; 7:511–519.

Simpson SL, Edwards LJ, Muller KE, Sen PK, Styner MA. A linear exponent AR(1) family of correlation structures. Statistics In Medicine. 2010; 29:1825–1838. [PubMed: 20658550]

Simpson SL, Edwards LJ, Muller KE, Styner MA. Kronecker Product Linear Exponent AR(1) Correlation Structures for Multivariate Repeated Measures. PLoS ONE. 2013; 9:e88864. [PubMed: 24586419]

Styner MA, Gerig G. Three-dimensional medial shape representation incorporating object variability. Computer Vision and Pattern Recognition CVPR. 2001:651–656.

Wishart J. The generalised product moment distribution in samples from a normal multivariate population. Biometrika. 1928; 20A:32–52.
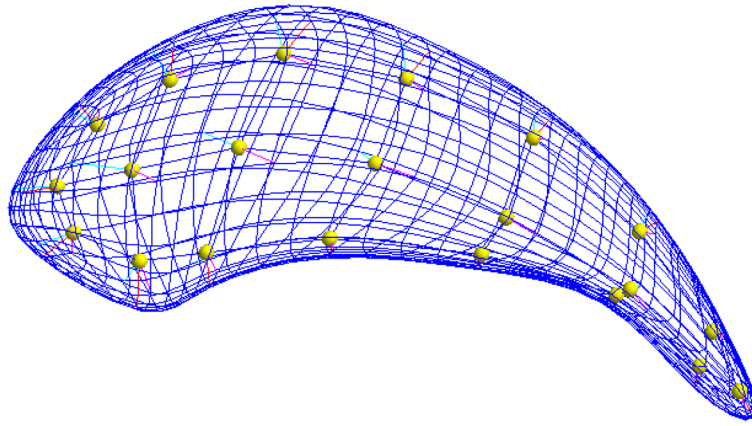
Figure 1.
M-rep shape representation model of the caudate.

Table 1

Simulated test size and power, $\alpha = 0.05$, 5,000 realizations.

		Structured Test				General Test			
		Test size × 100		Power × 100		Test size × 100		Power × 100	
N	# sig[1]	subset	full	subset	full	subset	full	subset	full
40	1	8.20	0.00	>99.99	>99.99	2.12	0.02	96.68	>99.99
	2	1.10	–	>99.99	–	0.12	–	77.46	–
	3	0.16	–	99.96	–	0.00	–	38.88	–
	4	0.02	–	87.56	–	0.00	–	9.04	–
80	1	7.50	0.14	>99.99	>99.99	2.30	0.24	>99.99	>99.99
	2	1.04	–	>99.99	–	0.08	–	99.82	–
	3	0.12	–	>99.99	–	0.00	–	90.96	–
	4	0.02	–	99.6	–	0.00	–	54.98	–
120	1	7.38	0.68	>99.99	>99.99	2.10	0.96	>99.99	>99.99
	2	0.90	–	>99.99	–	0.12	–	>99.99	–
	3	0.12	–	>99.99	–	0.02	–	99.44	–
	4	0.00	–	>99.99	–	0.02	–	84.60	–
160	1	8.12	1.26	>99.99	>99.99	2.04	1.68	>99.99	>99.99
	2	1.06	–	>99.99	–	0.06	–	>99.99	–
	3	0.06	–	>99.99	–	0.00	–	99.98	–
	4	0.00	–	>99.99	–	0.00	–	96.60	–
200	1	7.46	1.40	>99.99	>99.99	2.34	1.90	>99.99	>99.99
	2	0.98	–	>99.99	–	0.16	–	>99.99	–
	3	0.16	–	>99.99	–	0.00	–	>99.99	–
	4	0.02	–	>99.99	–	0.00	–	99.38	–

[1] # (out of $c = 4$) of significant p-values required for overall significance for the subset aLRT.

Table 2

Simulated test size and power by # (out of $c = 10$) of significant p-values required for overall significance, $\alpha = 0.05$, 5,000 realizations, $\max_i(s_i) = s = 21$, $\max_i(t_i) = 7$, $\mathrm{med}_i(t_i) = 3$.

N	# sig	Structured Test[1]		General Test[2]	
		Test size × 100	Power × 100	Test size × 100	Power × 100
140	1	99.16	> 99.99	96.74	> 99.99
	2	97.06	> 99.99	89.74	> 99.99
	3	91.24	> 99.99	76.62	> 99.99
	4	80.92	> 99.99	59.28	> 99.99
	5	65.70	> 99.99	39.34	> 99.99
	6	46.28	> 99.99	20.80	> 99.99
	7	28.12	> 99.99	9.46	> 99.99
	8	13.10	> 99.99	3.34	> 99.99
	9	4.32	> 99.99	0.70	99.98
	10	0.86	> 99.99	0.06	95.76
240	1	35.60	> 99.99	84.18	> 99.99
	2	13.80	> 99.99	65.92	> 99.99
	3	4.92	> 99.99	45.52	> 99.99
	4	1.40	> 99.99	28.24	> 99.99
	5	0.26	> 99.99	14.24	> 99.99
	6	0.02	> 99.99	5.92	> 99.99
	7	0.00	> 99.99	1.94	> 99.99
280	1	21.30	> 99.99	76.04	> 99.99
	2	5.52	> 99.99	54.80	> 99.99
	3	1.54	> 99.99	33.66	> 99.99
	4	0.36	> 99.99	17.50	> 99.99
	5	0.04	> 99.99	7.62	> 99.99
	6	0.02	> 99.99	3.02	> 99.99
	7	0.00	> 99.99	0.74	> 99.99
320	1	14.46	> 99.99	73.64	> 99.99
	2	3.00	> 99.99	49.60	> 99.99

N	# sig	Structured Test[1]		General Test[2]	
		Test size × 100	Power × 100	Test size × 100	Power × 100
3		0.56	> 99.99	29.22	> 99.99
4		0.10	> 99.99	14.42	> 99.99
5		0.02	> 99.99	6.04	> 99.99
6		0.00	> 99.99	1.76	> 99.99
7		0.00	> 99.99	0.38	> 99.99

[1] $v = 100$ (degrees of freedom) for each subset test (10 subset tests)

[2] $v = 75$ (degrees of freedom) for each subset test (10 subset tests)