University of Pennsylvania
UPenn Biostatistics Working Papers

Year 2016 *Paper* 47

Interpretable High-Dimensional Inference Via Score Maximization with an Application in Neuroimaging

Simon N. Vandekar* Philip T. Reiss†
Russell T. Shinohara‡

*University of Pennsylvania, simonv@mail.med.upenn.edu

†Department of Child and Adolescent Psychiatry, Department of Population Health, New York University, phil.reiss@nyumc.org

‡Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, rshi@upenn.edu

http://biostats.bepress.com/upennbiostat/art47

Interpretable High-Dimensional Inference Via Score Maximization with an Application in Neuroimaging

Simon N. Vandekar, Philip T. Reiss, and Russell T. Shinohara

Abstract

In the fields of neuroimaging and genetics a key goal is testing the association of a single outcome with a very high-dimensional imaging or genetic variable. Oftentimes summary measures of the high-dimensional variable are created to sequentially test and localize the association with the outcome. In some cases, the results for summary measures are significant, but subsequent tests used to localize differences are underpowered and do not identify regions associated with the outcome. We propose a generalization of Rao's score test based on maximizing the score statistic in a linear subspace of the parameter space. If the test rejects the null, then we provide methods to localize signal in the high-dimensional space by projecting the scores to the subspace where the score test was performed. This allows for inference in the high-dimensional space to be performed on the same degrees of freedom as the score test, effectively reducing the number of comparisons. We illustrate the method by analyzing a subset of the Alzheimer's Disease Neuroimaging Initiative dataset. Results suggest cortical thinning of the frontal and temporal lobes may be a useful biological marker of Alzheimer's risk. Simulation results demonstrate the test has competitive power relative to others commonly used.

Interpretable High-Dimensional Inference Via Score Maximization with an Application in Neuroimaging

Simon N. Vandekar, Philip T. Reiss, and Russell T. Shinohara*

May 11, 2016

Author's Footnote:

Simon N. Vandekar is Doctoral Candidate (email: simonv@mail.med.upenn.edu), Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104. Philip T. Reiss is Associate Professor (email: Phil.Reiss@nyumc.org), Department of Child and Adolescent Psychiatry, Department of Population Health at New York University, New York, NY 10016 and Department of Statistics University of Haifa, Haifa, Israel. Russell T. Shinohara is Assistant Professor of Biostatistics (email: rshi@mail.med.upenn.edu), Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104. SNV was supported by T32MH065218-11; PTR was suported by R01MH095836; RTS was supported by R01NS085211. The authors thank Wei Pan for helpful discussions related to this work.

*Corresponding author

Abstract

In the fields of neuroimaging and genetics a key goal is testing the association of a single outcome with a very high-dimensional imaging or genetic variable. Oftentimes summary measures of the high-dimensional variable are created to sequentially test and localize the association with the outcome. In some cases, the results for summary measures are significant, but subsequent tests used to localize differences are underpowered and do not identify regions associated with the outcome. We propose a generalization of Rao's score test based on maximizing the score statistic in a linear subspace of the parameter space. If the test rejects the null, then we provide methods to localize signal in the high-dimensional space by projecting the scores to the subspace where the score test was performed. This allows for inference in the high-dimensional space to be performed on the same degrees of freedom as the score test, effectively reducing the number of comparisons. We illustrate the method by analyzing a subset of the Alzheimer's Disease Neuroimaging Initiative dataset. Results suggest cortical thinning of the frontal and temporal lobes may be a useful biological marker of Alzheimers risk. Simulation results demonstrate the test has competitive power relative to others commonly used.
KEYWORDS: Score maximization, posthoc inference, association test, hypothesis test, neuroimaging

1 Introduction

Across scientific fields in which high-dimensional data are prominent, there is interest in testing the association of a single continuous or categorical outcome with a large number of independent variables. A common approach used in neuroimaging is to perform sequential tests to reduce the number of hypothesis tests. For example, in neuroimaging it is common to first perform a test for the association of a phenotype with an imaging variable averaged across the entire brain. If the test rejects the null hypothesis of no association between brain and phenotype, then subsequent tests are conducted on regional averages of the data, and then finally on every voxel in the image using multiplicity correction to address the thousands of tests performed in the image. Often, voxelwise results yield few or no significant findings due to reduced signal and the necessary adjustment for the large number of tests, even though the whole brain average data show a significant association.

In this paper we propose an approach to combine these series of tests into a single testing procedure. The framework is a generalization of Rao's score test when there are more independent variables than observations. Though the approach is designed to address hypothesis testing in neuroimaging, it is applicable in a wide range of scientific domains.

Score tests are appealing in neuroimaging and genetics, where the number of variables, p, is much larger than the sample size, n. When $p > n$ the score test has the advantage that it does not require an estimate of the model under the alternative hypothesis, thus has significantly reduced computation time relative to the Wald and likelihood ratio tests [Pan et al., 2014]. The sum of scores (Sum) test originally discussed by Rao [1948] has been used in genetics and neuroimaging [Pan, 2009, Kim et al., 2014].

To define the Sum test, assume Y_i iid observations with density f. Define the scores $\mathrm{U} = n^{-1}\sum_{i=1}^{n}\frac{\partial \log f(\mathrm{Y}_i|\theta)}{\partial\theta}(\theta_0)$ as the first derivative of the log likelihood evaluated at the null scaled by n^{-1}, and the Fisher information $\Omega(\theta_0) = \mathbb{E}\left\{[(\partial/\partial\theta)\log f(\mathrm{Y}_1 \mid \theta)]^T[(\partial/\partial\theta)\log f(\mathrm{Y}_1 \mid \theta)]\right\}$. The Sum test is based on the statistic

$$n\frac{(\mathrm{U}^T\zeta)^2}{\zeta^T\hat{\Omega}\zeta}, \tag{1}$$

where $\zeta \in \mathbb{R}^p$ is a known vector of weights. The denominator is an estimate of the variance of $\mathrm{U}^T\zeta$, so that the statistic is asymptotically χ^2_1 [Rao, 1948].

In genetics, the Sum test was introduced in the context of the model

$$h(\mathbb{E}\{\mathrm{Y}_i\}) = \alpha^T x_i + \beta^T g_i, \tag{2}$$

where $x_i \in \mathbb{R}^m$, is a vector of a small number ($m < n$) of observed covariates that includes a term for the intercept, $g_i \in \mathbb{R}^p$ is a vector of a large number of predictor variables of interest, α and β are unknown coefficient vectors, h is a known link function and Y is from an exponential family distribution. We seek to test $H_0 : \beta = 0$ versus the alternative $H_1 : \beta \neq 0$. The Sum test, in this framework, assumes that the components of β are equal, $\beta_1 = \beta_2 = \ldots = \beta_p = \beta_c$, and performs the test of $H_0 : \beta_c = 0$ [Pan, 2009]. In this case $\zeta = [1, \ldots, 1]$ in equation (1). More recent versions of the Sum test allow general weighting schemes [Madsen and Browning, 2009].

The Sum test is most powerful when the alternative differs from the null by small quantities [Rao, 1948]. However, the test has low power when there is a large number of variables that are not associated with the outcome [Pan et al., 2014]. This is due to the fact that the variance of the statistic in the numerator of (1) increases by adding nonassociated variables without increasing the expected value of the numerator. In addition, the test with positive weights has low power when there are coefficients with different signs. Another third limitation of the test is that it requires the *a priori* selection of weights; if multiple weight configurations are used then conservative multiple comparison corrections are necessary.

In the case of unknown weights when $p < n$, Rao [1948] proposed maximizing the sum test statistic with respect to the weights,

$$\max_{\zeta} n\frac{(\mathrm{U}^T\zeta)^2}{\zeta^T\hat{\Omega}\zeta} = n\mathrm{U}^T\hat{\Omega}^{-1}\mathrm{U}. \tag{3}$$

This statistic is χ^2_p under the null. This statistic cannot be used for high dimensional data due to the estimate $\hat{\Omega}$ being noninvertible when $p > n$. We refer to statistics of the type 3 as a MaxSum statistic because they maximizes the sum test with respect to the weight vector.

In this paper, we propose an extension of the MaxSum test for models with $p > n$. Our extension of the MaxSum can be thought of as a generalization of Rao's test in the case where the information matrix is noninvertible. The test has improved power over the Sum test, when $p > n$, by maximizing the statistic with respect to the vector ζ over a subspace, $\mathbb{L}$, of $\mathbb{R}^p$. The test works by rotating the scores for the original model to a lower dimensional

space where the information matrix is invertible. One advantage of the MaxSum is that the scores can be projected onto $\mathbb{L}$, and low rank multivariate inference on the projected scores can be performed to aid in localization of scores that contribute to the test statistic. We derive the asymptotic null distribution for the MaxSum statistic for models with likelihoods that satisfy some standard regularity conditions. For a normal linear model, we show how the finite sample distribution of the statistic can be calculated exactly.

A key advantage of the test is that it allows the investigator to incorporate known information about the structure of the signal, or to take advantage of the correlation structure in the data. In addition, the MaxSum can be used to combine score weights and naturally reduce the number of comparisons caused by testing multiple weights separately.

To demonstrate how the test can be used in neuroimaging we analyze data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. We investigate the association of cortical thickness with mild cognitive impairment (MCI). The outer surface of the brain (cortex) represents a highly folded sheet in 3-dimensional space. The thickness of the cortex is known to be affected in individuals with psychopathology and neurological illness. MCI is a subtle pre-Alzheimer's disease decline in cognitive functioning. There is interest in finding biological markers of MCI in order to identify those at risk for developing Alzheimer's disease. In this data set we seek to localize regions of the brain brain where cortical thinning provides additional information with regard to the diagnosis of MCI over what can be ascertained by neurocognitive scales alone.

In Section 2.1 we derive the test statistic and multivariate inference for likelihoods that satisfy general regularity conditions. In Section 2.2 we present a version of the test for normal data with a derivation of the finite sample distribution of the statistic. In Section 3 we discuss *post hoc* inference by projecting the rotated scores onto $\mathbb{L}$. Section 4 illustrates limitations of current methodology and demonstrates the MaxSum procedure in a neuroimaging data set from the Alzheimers Disease Neuroimaging Initiative (ADNI) data set where p =18,715 and $n = 628$. This analysis motivates the selection of the subspace $\mathbb{L}$. In Section 5 we compare the power of the MaxSum to several other tests used in genetics and neuroimaging.

For the remainder of the manuscript, we denote matrices by uppercase italic letters (X), vectors by lowercase (x), and random vectors by uppercase roman letters (X). Vector spaces are denoted with black-board letters ($\mathbb{X}$) and Greek letters are (usually unknown) model parameters. For the singular

value decomposition (SVD) of any matrix $X = UDV^T$, we will assume that the smallest dimensions of the matrices obtained are equal to the rank of the matrix X unless otherwise noted.

2 Maximization of the Sum Test

In this section we define the MaxSum statistic and give its asymptotic distribution for an arbitrary likelihood, given some regularity conditions. We discuss maximization of the sum test for normal linear models in Section 2.2, and then in Section 2.3 we discuss choices for the basis $\mathbb{L}$ in generalized linear models (GLMs).

2.1 Maximization of the Sum test for general distributions

We assume the observations Y_i, are independent with density $f(y \mid \theta)$ and we denote the log-likelihood $\ell(\theta) = \sum_{i=1}^{n} \log f(y_i \mid \theta)$. Consider a model of $m + p$ unknown parameters, $\theta = (\alpha, \beta)$, where $\beta \in \mathbb{R}^p$ are of interest and $\alpha \in \mathbb{R}^m$ are nuisance parameters. Under the null we assume $H_0 : \beta = \beta_0$. α is unrestricted and we estimate it with its MLEs. Let the information, $\Omega_F = \mathbb{E}\left[\{(\partial/\partial\theta) \log f(\mathrm{Y}_1 \mid \theta)\}^T \{(\partial/\partial\theta) \log f(\mathrm{Y}_1 \mid \theta)\}\right]$, for the full model be

$$\Omega_F(\theta) = \begin{bmatrix} \Omega_\alpha & \Omega_{\alpha\beta} \\ \Omega_{\beta\alpha} & \Omega_\beta \end{bmatrix}.$$

Then the efficient information for β is

$$\Omega(\theta) = \Omega_\beta - \Omega_{\beta\alpha}\Omega_\alpha^{-1}\Omega_{\alpha\beta}. \tag{4}$$

We will show that the asymptotic distribution of the MaxSum statistic can be found for any likelihood based on independent observations provided the same regularity conditions required for the convergence of the scores to a multivariate normal [Boos and Stefanski, 2013, p. 288]. First, we define the MaxSum statistic.

Definition 2.1. *Let $P_{\mathbb{L}}$ be the orthogonal projection matrix onto a linear space $\mathbb{L} \subset \mathbb{R}^p$ with $r = dim(\mathbb{L}) < n - m$. Let* $\mathrm{U} = n^{-1}(\partial/\partial\theta)\ell(\theta \mid \mathrm{Y})|_{\theta_0}$

denote the score vector for β *and* $\hat{\Omega}$ *be an estimate of the efficient information obtained from*

$$\hat{\Omega}_F = n^{-1} \sum_{i=1}^{n} \left(\frac{\partial \log f(\mathrm{Y}_i \mid \theta)}{\partial \theta^T} \right) \left(\frac{\partial \log f(\mathrm{Y}_i \mid \theta)}{\partial \theta} \right) \Big|_{\theta_0} \tag{5}$$

using the formula (4)*, where* $\hat{\theta}_0 = (\hat{\alpha}, \beta_0)$ *denotes the estimate of the parameter vector under the null hypothesis* $H_0 : \beta = \beta_0$*. Then the MaxSum statistic with respect to* $\mathbb{L}$ *is defined as*

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} = \max_{\zeta \in \mathbb{L}} n \frac{(\zeta^T \mathrm{U})^2}{\zeta^T \hat{\Omega}(\theta_0) \zeta} = \max_{\zeta \in \mathbb{R}^p} n \frac{(\zeta^T P \mathrm{U})^2}{\zeta^T P \hat{\Omega}(\theta_0) P \zeta}.$$

Under this definition we have the following theorem.

Theorem 2.2. *Assume all objects are as described in Definition 2.1. Let* $P_{\mathbb{L}} = QQ^T$ *where the columns of* Q *are any orthonormal basis for* $\mathbb{L}$*. Define*

$$V = V(\theta_0) = Q^T \Omega(\theta_0) Q,$$

and assume the estimate $\hat{\mathrm{V}} = Q^T \hat{\Omega}(\theta_0) Q$ *is invertible.*

In addition, assume the asymptotic normality of the scores, which depends on: (1) identifiability of the parameters, (2) the support of $f(y \mid \theta)$ *does not depend on* θ*, (3) The true parameter lies on the interior of the parameter space, (4)* $|(\partial^3 / \partial\theta_j \partial\theta_k \partial\theta_\ell) \log f(y \mid \theta)| \leq g(y)$*, with* $\mathbb{E}g(y) < \infty$*, and (5)* $\mathbb{E}\hat{\Omega}_F = \Omega_F(\theta_0)$*, for* $\hat{\Omega}_F$ *in Definition 2.1 and* Ω_F *is nonsingular and continuous with respect to* θ*. Then, the rotated scores are asymptotically normal*

$$n^{1/2} Q^T \mathrm{U} \xrightarrow{d} A_r(0, V). \tag{6}$$

and the MaxSum statistic is

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} = n \mathrm{U}^T Q \hat{\mathrm{V}}^{-1} Q^T \mathrm{U}, \tag{7}$$

with $\mathrm{R}^{\mathbb{L}}_{MaxSum} \xrightarrow{D} \chi^2_r$ *as* $n \to \infty$*.*

Theorem 2.2 requires that $\hat{\mathrm{V}}$ is nonsingular, however, in most cases it is possible to ensure that Q is in the column space of $\hat{\Omega}(\theta_0)$, so that $\hat{\mathrm{V}}^{-1}$ exists. The proof of Theorem 2.2 is given in A.1. We also demonstrate in Section A.1 that the result of Theorem 2.2 does not depend on the choice of Q. We show how $\mathbb{L}$ can be chosen for GLMs in Section 2.3 and for imaging data in the analysis of the ADNI dataset in Section 4.

2.2 Maximization of the Sum Test for Normal Linear Models

The finite-sample distribution for the MaxSum statistic can be found exactly for a normal linear model. Following the notation in equation (2), we define $X = [x_1, \ldots, x_n]^T$ to be an $n \times m$ full rank matrix of covariates for each observation, $\tilde{G} = [g_1, \ldots, g_n]^T$ to be an $n \times p$ full rank matrix of independent variables of interest with $p > n$, and $\tilde{\mathrm{Y}} = [\tilde{\mathrm{Y}}_1, \ldots, \tilde{\mathrm{Y}}_n]^T$ to be $n \times 1$ normal random vector with independent elements conditional on X and G. The Sum test with normal error, is based on the model

$$\tilde{\mathrm{Y}}_i = \alpha^T x_i + \beta^T g_i + \mathrm{E}_i,$$

where all variables are as previously defined and $\mathrm{E}_i \sim N(0, \sigma^2)$ are independent. If we let $AA^T = (I - H)$ be the SVD of the projection $(I - H)$, where $H = X(X^TX)^{-1}X^T$, and define $G = A^T\tilde{G}$ and $Y = A^T\tilde{\mathrm{Y}}$, then under the null $Y \sim N_{(n-m)}(0, \sigma^2 I)$.

The Sum test statistic for $H_0 : \beta = 0$ is [Rao, 1948, Lin and Tang, 2011]

$$\mathrm{R}_{Sum} = n\frac{(\mathrm{U}^T\zeta)^2}{\zeta^T\hat{\Omega}\zeta} = \frac{(\mathrm{Y}^TG\zeta)^2}{\hat{\sigma}^2 \times \zeta^TG^TG\zeta}, \tag{8}$$

where $\zeta \in \mathbb{R}^p$ is a known vector of weights, $\mathrm{U} = n^{-1}G^T\mathrm{Y}$ is the score vector evaluated at the MLEs under the null ($\mathbb{E}(\tilde{\mathrm{Y}}_i) = x_i\alpha$), and $\hat{\Omega} = n^{-1}\hat{\sigma}^2G^TG$ is an estimate of the efficient information matrix using (5).

The MaxSum statistic for this model with respect to a linear subspace $\mathbb{L}$ of $\mathbb{R}^p$ is defined from (8) as

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} = \max_{\zeta \in \mathbb{L}} \frac{(\mathrm{Y}^TG\zeta)^2}{\hat{\sigma}^2 \times \zeta^TG^TG\zeta}.$$

The following theorem gives a closed solution for $\mathrm{R}^{\mathbb{L}}_{MaxSum}$ as a function of Y, and gives the null distribution of the statistic.

Theorem 2.3. *Let $P_{\mathbb{L}}$ be the orthogonal projection matrix onto the space $\mathbb{L}$. Let W be a projection matrix onto the column space of GP. Then*

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} = (n - m)\frac{\mathrm{Y}^TW\mathrm{Y}}{\mathrm{Y}^T\mathrm{Y}}. \tag{9}$$

Also,

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} =_d \frac{r(n-m)}{r + (n-m-r)\mathrm{F}_{(n-m-r),r}},$$

where $\mathrm{F}_{(n-m-r),r}$ *is F-distributed with* $(n-m-r)$ *and* r *degrees of freedom.*

The proof can be found in appendix A.2. The form of equation (9) shows that for a normal linear model, the test statistic is a ratio of quadratic forms. Due to the rotation invariance of Y under the null, the finite-sample distribution of $\mathrm{R}^{\mathbb{L}}_{MaxSum}$ depends only on the sample size and the dimension of the basis. Thus, the choice of $\mathbb{L}$ does not affect the null distribution.

2.3 Specifying $\mathbb{L}$ in generalized linear models

Here we will discuss choices for the selection of $\mathbb{L}$ in the context of GLMs with the canonical link function. Define $\mathrm{Y} = [\mathrm{Y}_1, \ldots, \mathrm{Y}_n]^T$, $X = [x_1, \ldots, x_n]^T$, and $G = [g_1, \ldots, g_n]^T$ where objects are as defined in (2). For the GLM with canonical link, the scores are

$$\mathrm{U} = n^{-1} G^T (\mathrm{Y} - \hat{\mathrm{Y}}),$$

where $\hat{\mathrm{Y}} = [\hat{\mathrm{Y}}_1, \ldots \hat{\mathrm{Y}}_n]^T$ and $\hat{\mathrm{Y}}_i = h^{-1}(x_i^T \hat{\alpha})$ is the ith fitted value under the null. Let Γ be the diagonal matrix $\Gamma_{ii} = (\mathrm{Y}_i - \hat{\mathrm{Y}}_i)^2$. Then the efficient information estimate (5) is

$$\hat{\Omega} = n^{-1}\{G^T \Gamma G - G^T \Gamma X (X^T \Gamma X)^{-1} X^T \Gamma G\}$$

The score statistic is obtained from the scores and the estimated information as in expression (3).

If we condition on the design matrix G in a GLM, then the basis for $\mathbb{L}$ can be constructed from the principal components analysis (PCA) of G. We write the PCA of G in terms of the SVD $G = S_* D Q^T$, where the principal scores are $S = S_* D = GQ$.

With this basis the MaxSum test is equivalent to principal components regression. To see this, first note that principal component regression is defined by

$$h(\mathbb{E}(\mathrm{Y})) = X\alpha + S\beta_S.$$

The scores for β_S are

$$\mathrm{U}_S^T = n^{-1} Q^T G^T (\mathrm{Y} - \hat{\mathrm{Y}}) = Q^T \mathrm{U},$$

which are the same as the rotated scores in (6). The information estimate is

$$\begin{aligned}\hat{\Omega}_S &= n^{-1}\{S^T\Gamma S - S^T\Gamma X(X^T\Gamma X)^{-1}X^T\Gamma S\} \\ &= n^{-1}Q^T\{G^T\Gamma G - G^T\Gamma X(X^T\Gamma X)^{-1}X^T\Gamma G\}Q \\ &= Q^T\hat{\Omega}Q,\end{aligned}$$

which is the variance of the rotated scores. Thus the score test statistic, $n\mathrm{U}_S^T\hat{\Omega}_S^{-1}\mathrm{U}_S$, in principal component regression is equivalent the MaxSum statistic (7).

Another useful basis for $\mathbb{L}$ may be constructed from basis vectors that are indicators of variables that are expected to have a similar relationship with the outcome. The anatomical basis used in Section 4 is an example. For each basis vector q_j, $j = 1, \ldots, r$, we define a set $\mathcal{Q}_j \subset \{1, \ldots, p\}$ such that $\mathcal{Q}_j \cap \mathcal{Q}_{j'} = \varnothing$ and $q_{ij} = \mathbb{1}(i \in \mathcal{Q}_j)$. These define orthogonal basis vectors since the sets $\mathcal{Q}_j$ are disjoint. This basis is equivalent to averaging r subsets of the p independent variables and performing a hypothesis test of the regression onto the r averaged variables.

In general, for GLMs the rotation of the scores for the MaxSum test is equivalent to rotating the design matrix for the high-dimensional data to a lower dimensional space. An advantage of characterizing the test statistic in terms of a rotation is that *post hoc* inference of the contribution of the scores to the MaxSum statistic can be performed by considering the scores after projecting into $\mathbb{L}$.

3 *Post hoc* Inference to Localize Signal

If we reject the test of $H_0 : \beta = \beta_0$ using the MaxSum statistic it is of primary interest to investigate the contribution of the scores to the statistic. This can be done by projecting the rotated scores $Q^T\mathrm{U}$ back onto $\mathbb{L}$. Because the projected scores are distributed in a subspace of $\mathbb{R}^p$, inference is much less conservative compared to performing inference on the original score vector. The improved power of performing inference in the subspace is illustrated in Section 4.

Our aim is to construct a rejection region for the projected score vector under the null,

$$P\mathrm{U} \sim N(0, P\Omega P^T),$$

that has interpretable pointwise inference for each element $(P\mathrm{U})_j$. We begin by considering a box centered at the origin by finding $c \in \mathbb{R}^+$ such that

$$\mathbb{P}(|P\mathrm{U}|_\infty < c) = 1 - \alpha.$$

This set defines the region where the probability any element of the projected score vector is greater than c is α. We reject the null hypothesis at location j if the observed projected score $(Pu)_j > c$. One issue with this region is that the diagonal elements of the covariance of $P\mathrm{U}$ are not identical, so a symmetric box favors rejection for elements with larger variances.

To resolve this issue we scale by the inverse of the square root of the diagonal of the covariance of the projected scores, so that the variances for each element are identical. With this goal, we define the diagonal matrix $\Delta_{jj} = 1/\sqrt{\mathrm{Cov}(P\mathrm{U})_{jj}} = 1/\sqrt{(P\Omega P)_{jj}}$, then a more appropriate rejection region is defined by c that satisfies

$$\mathbb{P}(|\Delta P\mathrm{U}|_\infty < c) = 1 - \alpha. \tag{10}$$

By (6) we have

$$\Delta P\mathrm{U} \xrightarrow{d} \Delta Q V^{1/2}\mathrm{Z},$$

where $\mathrm{Z} \sim N_r(0, I)$. Thus we can approximate the region in (10) by finding c so that

$$\int_{-c \le \Delta Q V^{1/2} z \le c} \phi_r(z)dz = 1 - \alpha,$$

In practice we approximate this interval by plugging in estimates for Δ and $V^{1/2}$.

This integral is difficult to calculate due to the large dimensions of Q, but can be approximated quickly and easily using Monte Carlo simulations. B simulations are used to obtain an estimate the CDF, $\hat{F}_B(\cdot)$, which we use to obtain p-values for each observed projected score, $(Pu)_j$, by evaluating

$$p_{u_j} = 1 - \hat{F}_B(u_j), \tag{11}$$

or a rejection threshold can be obtained by using

$$q_{1-\alpha} = \hat{F}_B^{-1}(1 - \alpha). \tag{12}$$

The p-value for a given element of the observed projected score vector is the probability of observing a projected score as large as u_j under the global null $H_0 : \beta = \beta_0$.

It is important to note that the *post hoc* inference offers improved power by interpreting the projected scores. When the alternative hypothesis is true, the rejection regions for the projected scores do not necessarily control the type 1 error for the unprojected scores. This is demonstrated in the imaging simulations in Section 5.

4 ADNI Neuroimaging Data Analysis

We obtained data from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The ADNI is a longitudinal observational study designed to investigate the early biomarkers of Alzheimers disease. Mild cognitive impairment (MCI) represents a subtle pre-Alzheimer's Disease decline in cognitive performance. The goal of our analysis is to identify whether a subset of the neuroimaging data from the ADNI can provide more information regarding diagnosis of MCI than the standardized memory tests obtained as part of the ADNI. Moreover, we are interested in localizing areas of the cortex that differ between healthy controls (HC) and individuals with MCI. Three-dimensional T1-weighted structural images for 229 healthy controls and 399 subjects with MCI were obtained as part of the ADNI. This sample consists of subjects who had images a composite memory score available at baseline.

We perform the analysis in two ways: First, we proceed with standard software and analysis methods currently available for these type of neuroimaging data in open access software [Fischl, 2012]. Second, we use the MaxSum statistic and the high-dimensional inference procedure described above.

Cortical thickness was estimated using Freesurfer [Fischl and Dale, 2000, Dale et al., 1999]. Subjects' thickness data were registered to a standard template for analysis and smoothed at 10mm to reduce noise related to preprocessing and registration. The template contains 18,715 locations where cortical thickness is measured for each subject. Our goal is to identify whether the 18,715 cortical thickness measurements provide any additional information regarding the diagnosis of the individuals. For all analyses we include age, sex, and the composite memory score as covariates.

4.1 Standard Neuroimaging Analysis Procedure: Average and Vertex-wise Testing

Because neuroimaging studies typically collect many types of images with many covariates and possible outcomes, it is common to obtain a summary measure of a high-dimensional variable, and then proceed with further analysis if the summary measure appears to be associated with an endpoint of interest. In this analysis we first take the average of all the cortical thickness measurements across the cortical surface for each subject and perform a regression with diagnosis as the outcome using logistic regression. Specifically let C_i denote the average cortical thickness measurement for subject i, and X_i denote a vector with an intercept term, the age, an indicator for sex, and the composite memory score for subject i. Then we fit the model

$$\text{logit}\{\mathbb{P}(Y_i = 1 \mid C_i, X_i)\} = X_i^T \alpha + C_i \beta_C.$$

If there is a significant relationship with the average cortical thickness measurements, i.e. we reject $H_0 : \beta_C = 0$, then we will proceed by performing mass-univariate vertexwise analyses by running a separate model at each point on the cortical surface.

The analysis using the average cortical thickness variable suggests a highly significant association of cortical thickness with diagnosis, indicating that subjects with thinner cortices are more likely to have MCI (Table 1). Based on these results we decided to investigate the relationship at each vertex to localize where in the cortex the association occurs.

	Estimate (SE)	p-value
Age	-0.08 (0.02)	< 0.001
Sex (Male)	-0.37 (0.26)	0.15
Memory score	-3.22 (0.27)	< 0.001
Average cortical thickness	-4.23 (1.02)	< 0.001

Table 1: Results for the logistic regression of diagnosis onto covariates and whole-brain average cortical thickness. The memory score is a composite score included in the ADNI and was created using a factor analysis of well known memory assessments. Results for average cortical thickness indicate a highly significant association between cortical thickness and diagnosis. SE stands for standard error.

For the vertexwise analyses we use the software package Freesurfer to

perform FDR correction separately across each hemisphere (Figure 1 A). Due to the large number of tests (18,715 models) FDR-corrected results are spatially conservative and are difficult to interpret with the strength of the whole-brain average results. That is, given there is such a strong association between the average cortical thickness and diagnosis we expect more distributed associations than what can be seen in Figure 1 A. Uncorrected exploratory analyses were conducted to further identify regions related to the whole-brain results (Figure 1 B). The most significant results occur in left and right frontal lobes. These analyses suggest that thinning in larger portions of the frontal and temporal lobes is associated with increased risk of mild cognitive impairment, however these results were not found using a method that guarentees control of the familywise error rate.

4.2 MaxSum and High-Dimensional Inference Procedures

To use the MaxSum procedure we perform the following steps:

1. Select a subspace for $\mathbb{L}$;
2. Perform the MaxSum test for the association between the image and diagnosis;
3. If the test in step 2 rejects, then perform multivariate inference based on the projected scores in $\mathbb{R}^p$.

We select a basis for $\mathbb{L}$ in the two ways described in Section 2.3. First, we consider the basis constructed from the PCA of G after removing the effects of the covariates from G. For this analysis we chose the first $r = 20$ PCs as a basis for $\mathbb{L}$, however we also present results for the PCA truncated at several other dimensions ($r = 5, 10, 20, 50$) to demonstrate how the basis affects the results of the analysis. In addition we consider a basis, constructed from all regions of the anatomical atlas of Destrieux et al. [2010]. There are 74 regions in each hemisphere of the brain, so the rank of the basis is $r = 148$. If we were unwilling to condition on the covariance structure of G or the anatomical atlas, a basis could be constructed that approximates a predetermined covariance structure (e.g. a spatial AR(1)), or a covariance structure estimated from an independent sample.

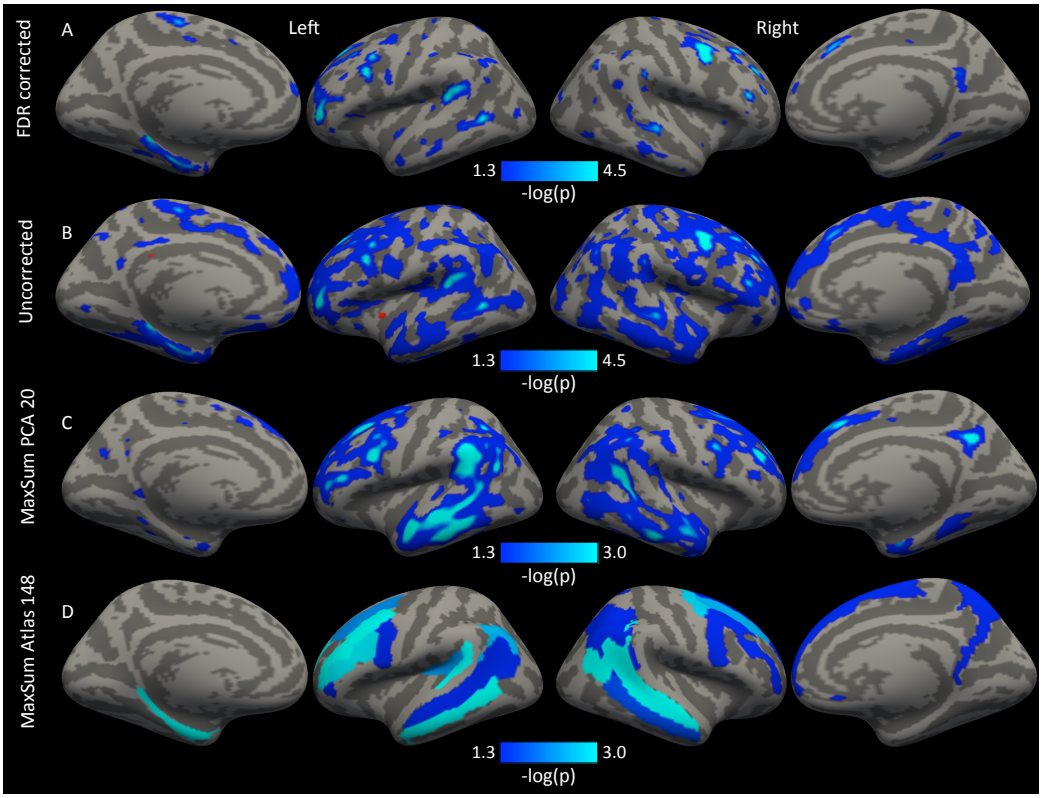

Figure 1: A comparison of inference procedures of the association between the imaging data and diagnosis. Color bars indicate $-\log_{10}(p)$. (A) FDR corrected vertexwise results, (B) Uncorrected vertexwise results, and (C & D) results based on MaxSum high-dimensional inference control the FWER of the projected scores. (C) The basis for $\mathbb{L}$ was selected from the first 20 components of the PCA of G. (D) The 148 dimensional basis constructed from the Destrieux anatomical atlas. Blue values have significant ($\alpha = 0.05$) negative association with diagnosis indicating that thinner cortex in these regions is associated with a MCI diagnosis.

Using the PCA basis, we reject the null hypothesis with the MaxSum test (Table 2), indicating that there is an association between the image and diagnosis conditioning on the effects of age, sex, and composite memory score. The test rejects at the $\alpha = 0.05$ threshold irrespective of which basis was used. Given the results of the MaxSum test we are then interested in investigating the regional variations in the scores.

Basis	r_{MaxSum}	p-value	95% threshold	99% threshold
PCA 5	24	< 0.001	3.0	3.6
PCA 10	41	< 0.001	3.4	4.0
PCA 20	50	< 0.001	3.7	4.2
PCA 50	91	< 0.001	4.0	4.4
Anatomical 148	179	0.04	3.5	3.9

Table 2: The χ^2 MaxSum statistic and associated p-values for the PCA basis with various basis dimensions: 5, 10, 20, and 50, and the anatomical atlas with 148 degrees of freedom, one for each label in the atlas. The degrees of freedom for the test are proportional to the dimension of the basis. The last two columns denote the 95% and 99% familywise error rejection thresholds for the standardized projected scores defined by the quantile function (12). The thresholds are obtained using 10,000 simulations. Note that increasing the basis dimension gives larger thresholds due to the increasing degrees of freedom. The projected scores are still highly correlated in the anatomical basis, so the rejection thresholds are lower.

To investigate the contributions of the scores to the MaxSum statistic we perform *post hoc* inference on the projected scores. As discussed in Section 3 we scale the projected scores by their variances in order to use constant rejection regions for easier interpretation. We use 10,000 simulations to obtain rejection regions for each of the basis dimensions. The simulations ran for all bases in less than 2 minutes.

Results suggest that thinner cortex in bilateral temporal and frontal lobes and right precuneus is associated with an increased risk of MCI (Figure 1 C & D). Results are given as $-\log_{10}(p)$ where p is obtained using the simulated distribution (11). These locations are known to be thinner in AD versus HC as well as in AD versus MCI [Singh et al., 2006] and the results here demonstrate that there are significant differences between MCI and HC in the same region. The results indicate that the degree of frontal

and temporal lobe thinning is correlated with diagnostic severity, and suggest that measurements of cortical thickness may provide useful information over neurospsychological scales in identifying people at risk for AD. Differences in these regions between MCI and HC were previously shown by Wang et al. [2009]; however the authors did not control for multiple comparisons or adjust for covariates.

To reiterate, the blue areas in Figure 1 C & D are based on low rank multivariate inference and control the FWER of the projected scores. The procedure demonstrates improved power over standard correction methods seen in Figure 1 A & B by performing inference in a lower dimensional space. The p-values obtained in Figures 1 and 2 use (11) and indicate the probability of observing a projected score statistic as extreme under the global null $H_0 : \beta = 0$. Though interpretation is restricted to the projected scores, the results align strongly with previous reports [Singh et al., 2006, Wang et al., 2009].

To demonstrate the implications of the choice of r on the inference of the projected scores we performed *post hoc* inference on the scores for 4 different PCA bases (Figure 2). It is clear from Figure 2 that increasing the dimension of the basis increases the spatial specificity of the results. However, the larger bases also come with the cost of a reduced power due to the larger degrees of freedom of the basis. This is also illustrated in Table 2, where the larger bases have a higher rejection threshold.

5 Neuroimaging Simulation Study

For this simulation study we perform analyses using data generated for the right hemisphere of the cortical thickness data from the ADNI dataset (p =9,361). The outcome of interest is categorical as in the ADNI analyses presented above. We select two anatomical regions (superior temporal sulcus and superior frontal sulcus) of 669 vertices total to have a negative association with the outcome and one region (anterior part of the cingulate gyrus and sulcus) of 191 vertices to have a positive association. To create a mean and covariance structure similar to real data within the regions of association, we create the mean vector and covariance matrices for the simulations from the full sample of subjects used in the ADNI Freesurfer analysis above, yielding two full rank covariance matrices, Σ_- and Σ_+ and mean vectors μ_+ and μ_-.

h!

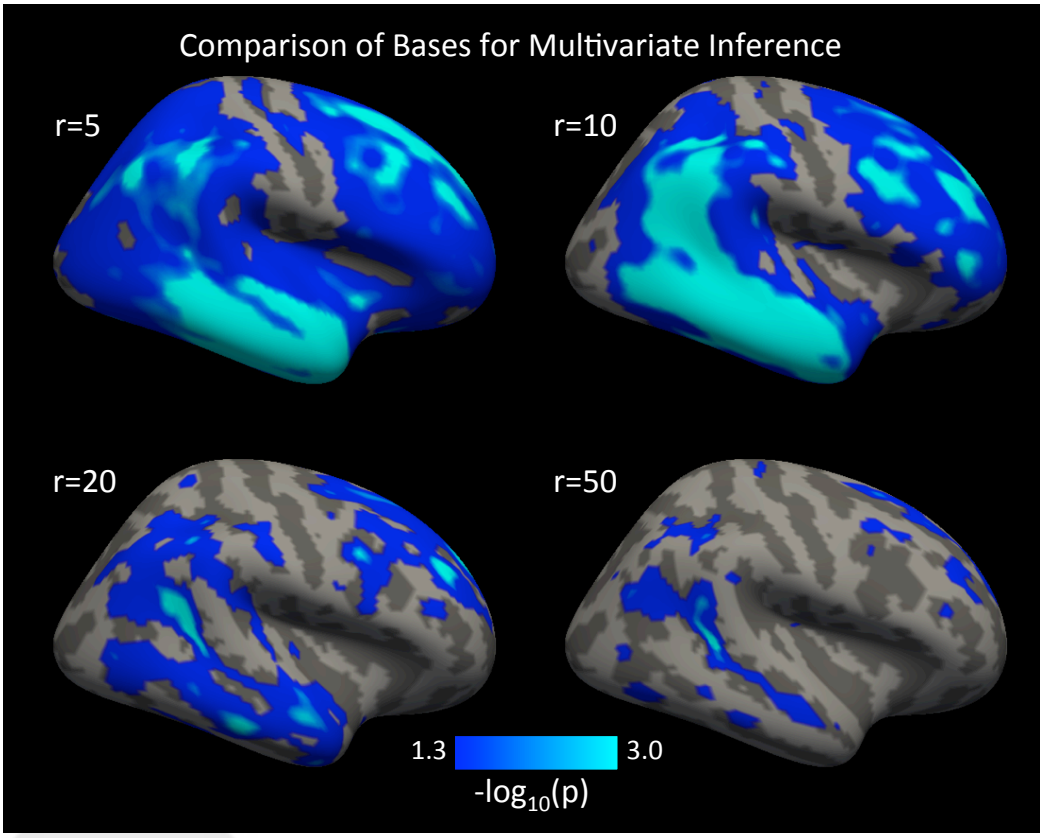

Figure 2: MaxSum high-dimensional inference for bases of various rank; 5, 10, 20, 50. Increasing the dimensionality of the basis increases spatial specificity, but comes with the cost of more conservative inference (see e.g. Table 2).

For each simulation we select a random subset of size n without replacement from the subset of control subjects used in the ADNI neuroimaging analysis. Data within the negatively and positively associated regions are generated as independent multivariate normal distributions for each subject, with covariance structures $G_{i,-} \sim N(\mu_+, \Sigma_-)$ and $G_{i,+} \sim N(\mu_-, \Sigma_+)$, respectively. We centered the imaging data prior to analysis.

In each simulation the outcome is generated under a logistic model

$$\text{logit}(\mathbb{E}Y_i) = \alpha_0 - \beta \mathbf{1}^T G_{i,-} + 2\beta \mathbf{1}^T G_{i,+}, \tag{13}$$

where α_0 is set to the log ratio of MCI to controls in the neuroimaging analyses section. $\mathbf{1}$, is a vector of ones, and β is an unknown parameter that we vary from 0 to 0.005. We multiply the values in the positive region by 2 to increase signal because it is a spatially smaller cluster than the two negative regions.

We construct the subspace $\mathbb{L}$ in two ways. The first basis is constructed in each sample from the first r principal components from a PCA of $G(I - H)$, where H is the projection onto the intercept. The second basis is constructed from regions of anatomical atlas of Destrieux et al. [2010], by randomly grouping the 74 regions into r groups and using normed indicator vectors for each group as the basis.

Denote the expectation of the score vector under the alternative by μ. If we denote the set of indices with a nonzero association with the outcome by J, then the expectation of the score vector is only nonzero for scores μ_j with $j \in J$. So, for indices with $j \notin J$ we report type 1 error and for indices with $j \in J$ we report power.

Similarly, the mean of the projected scores, $P\mu$, determines type 1 error and power for the projected scores PU. Because the basis is randomly constructed in each sample the nonzero elements of PU are different in each simulation. The FWER and FDR of the projected scores are reported for the basis constructed from the anatomical atlas. In general, no element in the projected scores based on PCA construction is exactly zero, so type 1 error is not reported for this basis.

We perform 1000 simulations for sample sizes of $n = 100, 200$ and compare the MaxSum test for bases with dimensions of $r = 5, 10, 20, 50$. In addition, we compare the MaxSum test to the sequence kernel association test (SKAT) [Wu et al., 2011], the sum of powered scores (SPU) test using the infinity norm, which corresponds to testing the max across the scores [Pan

et al., 2014], and the adaptive sum of powered scores test (aSPU), which has competitive power than many other score tests [Pan et al., 2014]. The SKAT is known to be more powerful if there is a distributed signal, and the SPU infinity norm will be most powerful for sparse signal. The aSPU test combines multiple tests by choosing one with the smallest p-value amongst a set of powered score tests. Permutation tests are used to assess the significance for these statistics. We assess pointwise power and type 1 error of the MaxSum multivariate inference with uncorrected, Bonferroni corrected, and FDR corrected results. We also compare FWER and FDR between methods. For these comparisons we assess the type 1 error for the unprojected scores using inference designed for the projected scores.

The MaxSum test with both bases demonstrates superior power to the other tests (Figure 3), due to its ability to remove the influence of unassociated scores from the test by maximizing over the basis. In addition, the MaxSum test with the PCA takes advantage of the covariance structure of the scores and the spatial information of the signal. aSPU is adaptive to the sparsity of the signal, so performs better than the SKAT, but does not use the information in the covariance of the scores to leverage power. Note that the design matrix being random does not affect the Type 1 error even though the bases are constructed conditionally on the design matrix. This justifies the use of the test in a wide range of fields where the independent variables are themselves random.

As expected, the MaxSum *post hoc* inference procedure controls the FWER of the projected scores for all basis dimensions (Table 3). It is not possible to assess type 1 error for the PCA basis as no projected score for this basis is exactly zero after projection.

	Anatomical 5	Anatomical 10	Anatomical 20	Anatomical 50
FWER	0.03	0.04	0.04	0.04
FDR	0.02	0.03	0.02	0.02

Table 3: Error rates for the projected scores for the anatomical bases of dimension 5, 10, 20 and 50.

In general, the MaxSum multivariate inference procedure does not control the FWER or FDR of the unprojected scores (Table 4) as the inference is intended for the projected scores. However, for larger PCA bases our procedure does control the FDR (bold rows in Table 4). This is likely because

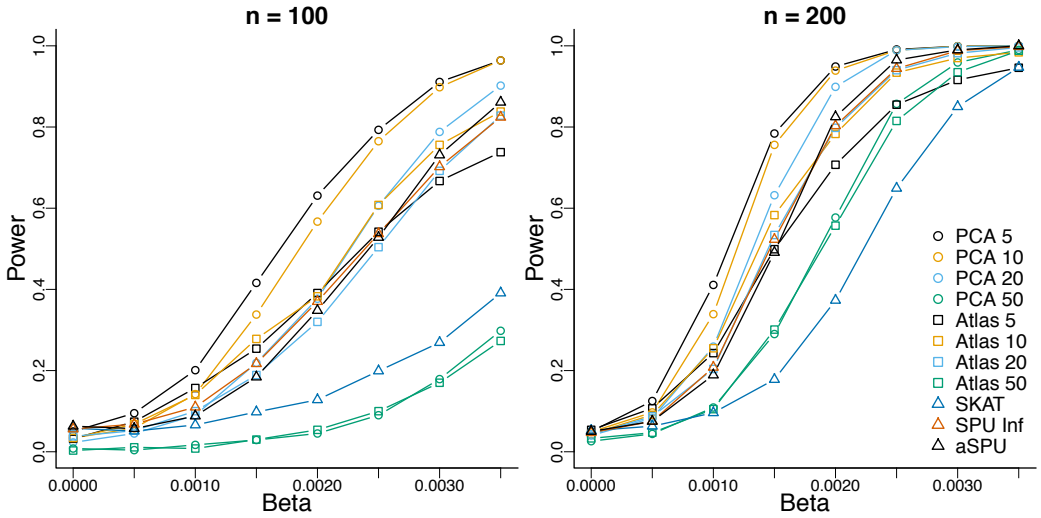

Figure 3: Power results for the MaxSum test with various bases compared to aSPU and the SKAT. "PCA r" indicates the basis formed from the first r components of the PCA of the design matrix G. "Atlas r" indicates the bases formed from the anatomical atlas.

the projection captures most of the signal in the expected value of the score vector, μ, so that the projection $P\mu$ is close to μ. In practice, it is not possible to tell if inference for the projected scores will control the type 1 error for the unprojected score vector.

The vertexwise error rate for the MaxSum inference procedure for the unprojected scores using the PCA 10 basis is low while maintaining better vertexwise power than FDR (Figure 4; PCA 10). This is because in any given sample there may be a high FWER, but the errors across samples do not appear in the same locations. The Benjamini-Hochberg and Bonferroni corrections both work well at controlling the vertexwise type 1 error rate however have lower power compared to the PCA-based MaxSum procedure (Figure 4). The bases constructed from the anatomical atlas tend to have large regions of vertexwise type 1 error for the unprojected scores. At the largest basis dimension (50) the atlas allows for enough specificity to reduce the vertexwise error. All methods have lower power to detect the positive cluster than the two negative clusters. This is possibly due to the characteristics of the covariance structure in the positive cluster which overlaps gyral and sulcal regions.

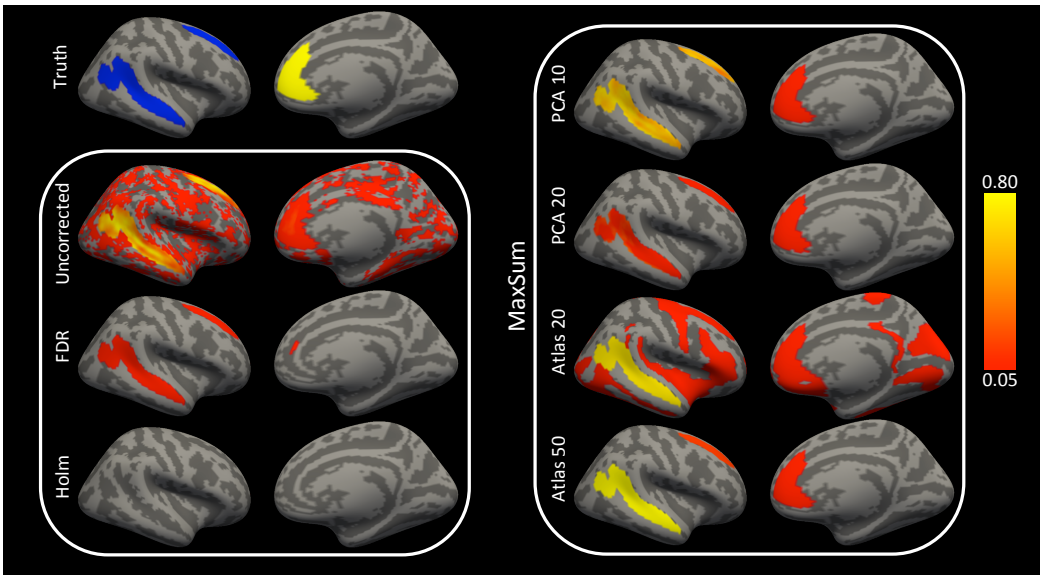

Figure 4: Vertexwise power or type 1 error is measured as the proportion of simulated samples where each of the testing procedures rejects the null at a given location. Results are shown for $\beta = 0.002$. "Truth" indicates locations where signal was simulated according to the model (13).

	HR	FDR	FWE
PCA 5	0.72	0.12	0.72
PCA 10	**0.58**	**0.04**	**0.48**
PCA 20	**0.37**	**0.01**	**0.22**
PCA 50	**0.13**	**0.01**	**0.08**
Uncorrected	0.62	0.43	1.00
Holm	**0.01**	**0.01**	**0.01**
BH	**0.12**	**0.05**	**0.25**
Anatomical 5	0.16	0.79	0.25
Anatomical 10	0.35	0.64	0.58
Anatomical 20	0.53	0.44	0.84
Anatomical 50	0.64	0.17	0.71

Table 4: Hit (HR), false discovery (FDR), and family-wise error (FWER) rates for the unprojected scores. Higher dimensions of the PCA basis control the FDR while maintaining a higher hit rate for the unprojected scores. Note, however, that in practice it is not possible to tell what dimension of the basis is required to control the FDR of the unprojected scores. Bold rows control the FDR at $q = 0.05$.

6 Discussion

We proposed the MaxSum test, which maximizes the weights for the Sum statistic in a subspace of the parameter space. The procedure offers powerful posthoc inference on the projected scores by performing inference in the subspace of where the test statistic was estimated. Because the posthoc inference is based on the same model and degrees of freedom as the MaxSum statistic, the interpretation of high-dimensional results agree closely with the results from the MaxSum test.

Instead of choosing a specific value for the weight vector, ζ, our methodology allows the investigator to select a space to consider for ζ. The ability to choose a space makes the procedure very flexible. For example, in imaging the basis for the space can be chosen based on anatomical or functional labels, or from data acquired in another imaging modality. Particular hypotheses can be targeted by selecting a basis that contains vectors that are indicators of certain regions or contain variable weights that capture specific spatial patterns. If orthogonal indicator vectors are used as the basis, then the approach can be seen as testing averages of subregions of the data as in

Section 4. In this case, the MaxSum procedure can be seen as a multiplicity correction procedure that accounts for the correlation structure of the tests.

The methodology inspires several theoretical questions. Further research could investigate how this approach relates to regularization in parameter estimation and whether regularization could be applied to the scores to produce similar results to maximizing the score test in a subspace of $\mathbb{R}^p$. The rotated scores correspond to a reparameterization of the original likelihood with fewer parameters. It might be possible to develop theory that allows for more general dimension-reducing transformations of the parameters where inference in the higher dimensional parameter space can be performed by mapping back up to a subspace of the original parameter space, although this procedure may reduce interpretability. We showed in simulations that larger basis dimensions of the PCA basis control the FDR of the unprojected scores. It will be useful to study under what conditions inference based on the projected scores will control the FDR of the unprojected scores. The MaxSum procedure allows for inference of the projected scores. This requires that the projection of the scores be something that is interpretable in the scientific field.

A. APPENDIX

A.1. Proof of Theorem 2.2

The following Lemma is used to guarantee that Ω is invertible given our assumption that Ω_F is invertible. The theorem is given here without proof. The proof can be found in Gallier [2001].

Lemma A.1. *Let*

$$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$$

be an $m \times m$ invertible symmetric matrix and A and C be $p \times p$, and $q \times q$ principal submatrices of M, Then A and C are invertible and the Schur complement of C, defined by $S = A - BC^{-1}B^T$, is invertible.

Proof. of Theorem 2.2 Let $\phi = Q^T\zeta$. Then the MaxSum statistic is

$$\begin{aligned} \mathrm{R}^{\mathbb{L}}_{MaxSum} &= \max_{\zeta\in\mathbb{R}^p} n\frac{(\zeta^T P\mathrm{U})^2}{\zeta^T P\hat{\Omega}P\zeta} \\ &= \max_{\phi\in\mathbb{R}^r} n\frac{(\phi^T Q^T\mathrm{U})^2}{\phi^T\hat{\mathrm{V}}\phi}. \end{aligned} \tag{14}$$

Under the null, the numerator of (14) is asymptotically normal (6), which follows by the result for the scores and the fact that linear combinations of normal random variables are normal [Boos and Stefanski, 2013].

$\hat{\mathrm{V}}$ is positive definite by assumption, so there exists a symmetric invertible matrix $\hat{V}^{1/2}$ that satisfies $\hat{\mathrm{V}}^{1/2}\hat{\mathrm{V}}^{1/2} = \hat{\mathrm{V}}$. Let $\psi = \hat{\mathrm{V}}^{1/2}\phi$, then we see that

$$\begin{aligned} \mathrm{R}^{\mathbb{L}}_{MaxSum} &= \max_{\psi\in\mathbb{R}^r} n\frac{(\psi^T\hat{\mathrm{V}}^{-1/2}Q^T\mathrm{U})^2}{\psi^T\psi} \\ &= \max_{\psi\in\mathbb{R}^r} n\frac{(\psi^T\hat{\mathrm{V}}^{-1/2}Q^T\mathrm{U})(\mathrm{U}QV^{-1/2}\psi)}{\psi^T\psi} \\ &= n\times \mathrm{eigmax}\{\hat{\mathrm{V}}^{-1/2}Q^T\mathrm{U}\mathrm{U}^TQ\hat{\mathrm{V}}^{-1/2}\} \\ &= n\times \mathrm{eigmax}\{\mathrm{U}^TQ\hat{\mathrm{V}}^{-1/2}\hat{\mathrm{V}}^{-1/2}Q^T\mathrm{U}\} \\ &= n\mathrm{U}^TQ\hat{\mathrm{V}}^{-1}Q^T\mathrm{U} \end{aligned}$$

The second line is a Rayleigh quotient and maximization with respect to ψ corresponds to the largest eigenvalue of $\hat{\mathrm{V}}^{-1/2}Q^T\mathrm{U}\mathrm{U}^TQ\hat{\mathrm{V}}^{-1/2}$. eigmax$\{\cdot\}$ denote the function that returns the largest eigenvalue of a matrix.

By (6), $n^{1/2}V^{-1/2}Q^T\mathrm{U} \xrightarrow{d} N_r(0, I)$. $V^{-1/2}$ exists given our assumptions in Theorem 2.2: The assumption about the invertibility of Ω_F guarantees Ω is invertible by Lemma A.1, because Ω is the Schur complement of Ω_α. This implies Ω is full rank, so $V = Q^T\Omega Q$ is invertible.

The variance estimate of $Q^T\mathrm{U}$ is

$$\begin{aligned} \hat{\mathrm{V}}(\theta_0) &= Q^T\hat{\Omega}(\theta_0)Q \\ &= (Q^T\hat{\Omega}_\beta Q - Q^T\hat{\Omega}_{\beta\alpha}\hat{\Omega}_\alpha^{-1}\hat{\Omega}_{\alpha\beta}Q). \end{aligned}$$

Convergence of $\hat{\mathrm{V}}(\theta_0) \xrightarrow{p} V(\theta_0)$ follows from the continuous mapping theorem and the convergence of its components by the weak law of large numbers, e.g. $Q^T\hat{\Omega}_\beta Q \xrightarrow{p} Q^T\Omega_\beta Q$, and $Q^T\hat{\Omega}_{\beta\alpha} \xrightarrow{p} Q^T\Omega_{\beta\alpha}$. Then, since the matrix inverse is continuous, $\hat{\mathrm{V}}(\theta_0)^{-1/2} \xrightarrow{p} V^{-1/2}$ by the continuous mapping theorem.

Importantly, for $r < n < p$, we necessarily have $\mathbb{P}(\{\hat{\Omega}(\theta_0) \text{ is invertible}\}) = 0$, however we are able to assume $\mathbb{P}(\{\hat{\mathrm{V}}(\theta_0) \text{ is invertible}\}) = 1$.

Finally, Slutsky's theorem gives the convergence in distribution of

$$n^{1/2}(\hat{\mathrm{V}}^{-1/2}V^{1/2})V^{-1/2}Q^T\mathrm{U} \xrightarrow{d} \mathrm{Z} \sim N_r(0, I).$$

So, $\mathrm{R}^{\mathbb{L}}_{MaxSum} \xrightarrow{d} \mathrm{Z}^T\mathrm{Z} \sim \chi^2_r$.

□

Remark. The conclusion of Theorem 2.2 implies that expression 7 does not depend on the choice of Q. This fact can also be shown directly, as follows. Consider another matrix Q_* with orthonormal columns such that $P = Q_*Q_*^T$, and accordingly define $\hat{V}_* = Q_*^T\hat{\Omega}(\theta_0)Q_*$. Then $Q_* = QM$ where $M = Q^TQ_*$. Since $P = QMQ_*^T$ is of rank r, M is of rank r and hence invertible, so

$$Q_*\hat{V}_*^{-1}Q_*^T = QM(M^T\hat{V}M)^{-1}M^TQ^T = Q\hat{V}^{-1}Q^T,$$

and thus formula (7) for the MaxSum statistic is unchanged by substituting $Q_*, \hat{V}_*$ for $Q, \hat{V}$.

A.2. Proof of Theorem 2.3

Proof. We will ignore the term $\frac{1}{\hat{\sigma}^2}$ in the maximization as it is constant with respect to ζ. Define $\mathbb{M} = G\mathbb{L}$ to be the column space of W. From the definition of $\mathrm{R}^{\mathbb{L}}_{MaxSum}$

$$\begin{aligned}
\mathrm{R}^{\mathbb{L}}_{MaxSum} &\propto \max_{\zeta\in\mathbb{R}^{\mathrm{l}}} \frac{(\mathrm{Y}^TGP\zeta)^2}{\zeta^TPG^TGP\zeta} \\
&= \max_{\zeta\in\mathbb{R}^{\mathrm{l}}} \frac{(\mathrm{Y}^TWGP\zeta)^2}{\zeta^TPG^TWGP\zeta} \\
&= \max_{\gamma\in\mathbb{M}} \frac{(\mathrm{Y}^TW\gamma)^2}{\gamma^TW\gamma}, \qquad (15)
\end{aligned}$$

where $\gamma = GP\zeta$. Because γ must be in the subspace $\mathbb{M}$ since it is the column space of W and GP. The solution to the Rayleigh quotient (15) is the solution to the largest generalized eigenvalue problem,

$$W\mathrm{Y}\mathrm{Y}^TW\gamma = \lambda_{max}W\gamma,$$

where $\lambda_{max} \in \mathbb{R}$, and $\|\gamma\| = 1$. By letting $\phi = W\gamma$ the solution is equivalent to the largest eigenvalue problem

$$W\mathrm{Y}\mathrm{Y}^T W\phi = \lambda_{max}\phi$$

Then

$$\mathrm{eigmax}(W\mathrm{Y}\mathrm{Y}^T W) = \mathrm{eigmax}(\mathrm{Y}^T W\mathrm{Y}) = \mathrm{Y}^T W\mathrm{Y}.$$

Finally, we have

$$\mathrm{R}^{\mathbb{L}}_{MaxSum} = (n-m)\frac{\mathrm{Y}^T W\mathrm{Y}}{\mathrm{Y}^T\mathrm{Y}}.$$

$\mathrm{R}^{\mathbb{L}}_{MaxSum}$ is a ratio of quadratic forms, but can be expressed in terms of a variable with an F-distribution.

To see this, first note that

$$\frac{\mathrm{Y}^T T}{\mathrm{Y}^T W\mathrm{Y}} = 1 + \frac{\mathrm{Y}^T(I-W)\mathrm{Y}}{\mathrm{Y}^T W\mathrm{Y}}. \tag{16}$$

The numerator and denominator of the random term on the left hand side are independent since $P(I-P) = 0$. So (16) is distributed as $1+\frac{n-m-r}{r}F_{(n-m-r),r}$. Thus

$$\begin{aligned}\mathrm{R}^{\mathbb{L}}_{MaxSum} &=_d \frac{(n-m)}{1+\frac{(n-m-r)}{r}F_{(n-m-r),r}} \\ &= \frac{r(n-m)}{r+(n-m-r)F_{(n-m-r),r}}.\end{aligned}$$

□

References

Dennis D Boos and L. A Stefanski. *Essential Statistical Inference*. Springer, New York, NY, 2013. ISBN 978-1-4614-4817-4 978-1-4614-4818-1. URL http://link.springer.com/10.1007/978-1-4614-4818-1.

Anders M. Dale, Bruce Fischl, and Martin I. Sereno. Cortical surface-based analysis: I. Segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999. URL http://www.sciencedirect.com/science/article/pii/S1053811998903950.

Christophe Destrieux, Bruce Fischl, Anders Dale, and Eric Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1):1–15, October 2010. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2010.06.010. URL http://www.sciencedirect.com/science/article/pii/S1053811910008542.

Bruce Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, August 2012. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2012.01.021. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3685476/.

Bruce Fischl and Anders M. Dale. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055, September 2000. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.200033797. URL http://www.pnas.org/content/97/20/11050.

Jean H. Gallier. *Geometric Methods and Applications: For Computer Science and Engineering*. Springer Science & Business Media, 2001. ISBN 978-0-387-95044-0.

Junghi Kim, Jeffrey R. Wozniak, Bryon A. Mueller, Xiaotong Shen, and Wei Pan. Comparison of statistical tests for group differences in brain functional networks. *NeuroImage*, 101:681–694, November 2014. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2014.07.031. URL http://www.sciencedirect.com/science/article/pii/S1053811914006144.

Dan-Yu Lin and Zheng-Zheng Tang. A General Framework for Detecting Disease Associations with Rare Variants in Sequencing Studies. *American Journal of Human Genetics*, 89(3):354–367, September 2011. ISSN 0002-9297. doi: 10.1016/j.ajhg.2011.07.015. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3169821/.

Bo Eskerod Madsen and Sharon R. Browning. A groupwise association test for rare mutations using a weighted sum statistic. *PLoS genetics*, 5(2):e1000384, February 2009. ISSN 1553-7404. doi: 10.1371/journal.pgen.1000384.

Wei Pan. Asymptotic Tests of Association with Multiple SNPs in Linkage Disequilibrium. *Genetic epidemiology*, 33(6):497–507, September 2009. ISSN 0741-0395. doi: 10.1002/gepi.20402. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2732754/.

Wei Pan, Junghi Kim, Yiwei Zhang, Xiaotong Shen, and Peng Wei. A Powerful and Adaptive Association Test for Rare Variants. *Genetics*, 197(4): 1081–1095, August 2014. ISSN 0016-6731, 1943-2631. doi: 10.1534/genetics.114.165035. URL http://www.genetics.org/content/197/4/1081.

C. Radhakrishna Rao. Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 44, pages 50–57. Cambridge Univ Press, 1948.

Vivek Singh, Howard Chertkow, Jason P. Lerch, Alan C. Evans, Adrienne E. Dorr, and Noor Jehan Kabani. Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimer's disease. *Brain*, 129(11):2885–2893, 2006. URL https://brain.oxfordjournals.org/content/129/11/2885.

Liya Wang, Felicia C. Goldstein, Emir Veledar, Allan I. Levey, James J. Lah, Carolyn C. Meltzer, Chad A. Holder, and Hui Mao. Alterations in cortical thickness and white matter integrity in mild cognitive impairment measured by whole-brain cortical thickness mapping and diffusion tensor imaging. *American Journal of Neuroradiology*, 30(5):893–899, 2009. URL https://www.ajnr.org/content/30/5/893.full.

Michael C. Wu, Seunggeun Lee, Tianxi Cai, Yun Li, Michael Boehnke, and Xihong Lin. Rare-Variant Association Testing for Sequencing Data with the Sequence Kernel Association Test. *American Journal of Human Genetics*, 89(1):82–93, July 2011. ISSN 0002-9297. doi: 10.1016/j.ajhg.2011.05.029. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3135811/.