Generalized reduced rank latent factor regression for high dimensional tensor fields, and neuroimaging-genetic applications

Chenyang Tao[a,b], Thomas E. Nichols[c], Xue Hua[d], Christopher R. K. Ching[d,e], Edmund T. Rolls[b,f], Paul Thompson[d,g], Jianfeng Feng[a,b,*]

and the Alzheimer's Disease Neuroimaging Initiative[†]

[a] Centre for Computational Systems Biology and School of Mathematical Sciences, Fudan University, Shanghai, PR China [b] Department of Computer Science, University of Warwick, Coventry, UK [c] Department of Statistics, University of Warwick, Coventry, UK [d] Imaging Genetics Center, Institute for Neuroimaging & Informatics, University of Southern California, Los Angeles, CA, USA [e] Interdepartmental Neuroscience Graduate Program, UCLA School of Medicine, Los Angeles, CA, USA [f] Oxford Centre for Computational Neuroscience, Oxford, UK [g] Departments of Neurology, Psychiatry, Radiology, Engineering, Pediatrics, and Ophthalmology, USC, Los Angeles, CA, USA

[*]Address for correspondence: Jianfeng Feng, Centre for Computational Systems Biology, Fudan University, 220 Handan Road, 200433, Shanghai, PRC.
E-mail: jianfeng64@gmail.com

[†] Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Summary. We propose a generalized reduced rank latent factor regression model (GRRLF) for the analysis of tensor field responses and high dimensional covariates. The model is motivated by the need from imaging-genetic studies to identify genetic variants that are associated with brain imaging phenotypes, often in the form of high dimensional tensor fields. GRRLF identifies from the structure in the data the effective dimensionality of the data, and then jointly performs dimension reduction of the covariates, dynamic identification of latent factors, and nonparametric estimation of both covariate and latent response field. After accounting for the latent and covariate effect, GRLLF performs a nonparametric test of the remaining factor of interest. GRRLF provides a better factorization of the signals compared with common solutions, and is less susceptible to overfitting because it exploits the effective dimensionality. The generality and flexibility of GRRLF also allow various statistical models to be handled in a unified framework and solutions can be efficiently computed. Within the field of neuroimaging, it improves the sensitivity for weak signals and is a promising alternative to existing approaches. The operation of the framework is demonstrated with both synthetic datasets and a real-world neuroimaging example in which the effects of a set of genes on the structure of the brain at the voxel level were measured, and the results compared favorably with those from existing approaches.

KEY WORDS: *Dimension reduction; Generalized linear model; High dimensional tensor field; Latent factor; Least squares kernel machines; Nuclear norm regularization; Reduced rank regression; Riemannian manifold;*

1 Introduction

The past decade has witnessed the dawn of the big data era. Advances in technologies in areas such as genomics and medical imaging, amongst others, have presented us with an unprecedentedly large volume of data characterized by high dimensionality. This not only brings opportunities but also poses new challenges to scientific research. Neuroimaging-genetics, one of the burgeoning interdisciplinary fields emerging in this new era, aims at understanding how the genetic makeup affects the structure and function of the human brain and has received increasing interest in recent years.

Starting with candidate gene and candidate phenotype studies, imaging-genetic methods have made significant progress over the years (Thompson et al., 2013; Liu and Calhoun, 2014; Poline et al., 2015). Different strategies have been implemented to combine the genetic and neuroimaging information, producing many promising results (Hibar et al., 2015; Richiardi et al., 2015; Jia et al., 2016). Using a few summary variables of the brain features is the most popular approach in the literature (Joyner et al., 2009; Potkin et al., 2009; Vounou et al., 2010); voxel-wise and genome-wide association approaches offer a more holistic perspective and are used in exploratory studies (Hibar et al., 2011; Vounou et al., 2012); multivariate analyses have also been used to capture the epistatic and pleiotropic interactions, therefore boosting the overall sensitivity (Hardoon et al., 2009; Ge et al., 2015a,b). Apart from the population studies, family-based studies offer additional insights on the genetic heritability (Ganjgahi et al., 2015). Recently, a few probabilistic approaches have been proposed to jointly model the interactions between genetic factors, brain endophenotypes and behavior phenotypes (Batmanghelich et al., 2013; Stingo et al., 2013), and some Bayesian methods originally developed for eQTL studies can also be applied to imaging-genetic problems (Zhang and Liu, 2007; Jiang and Liu, 2015).

The trend in imaging-genetics is to embrace brain-wide genome-wide association studies with multivariate predictors and responses, but this is challenged by the combinatorial complexity of the problem. For example, the probabilistic formulations do not scale well with dimensionality; and standard brute force massive univariate approaches (Stein et al., 2010a; Vounou et al., 2012)

treat each voxel and predictor as independent units and compute pairwise significance, and the loss of spatial information and the colossal multiple comparison corrections involved have high costs in terms of sensitivity (Hua et al., 2015). Various attempts have been made to remedy this. Some approaches involve dimension reduction techniques, which either first embed genetic factors onto some lower dimensional space using methods such as principal component analysis (PCA) before subsequent analyses (Hibar et al., 2011), or jointly project genetic factors and imaging traits by methods such as parallel independent component analysis (ICA), canonical correlation analysis (CCA) and partial least square (PLS) (Liu et al., 2009; Le Floch et al., 2012, 2013). These methods often lack model interpretability. Other popular approaches enforce penalties or constraints to regularize the solutions, for example (group) sparsity or rank constraints (Wang et al., 2012a,b; Vounou et al., 2012; Lin et al., 2015; Huang et al., 2015). But they are usually difficult to compute and the significance of the findings can not be directly evaluated.

One path towards more efficient estimation for brain-wide association, both in the statistical and computational sense, is to exploit the inherent spatial structure from the neuroimaging data. Two prominent examples in this direction are *random field theory* based methods (Worsley et al., 1996; Penny et al., 2011; Ge et al., 2012) and *functional* based methods (Wahba, 1990; Ramsay and Silverman, 2005; Reiss and Ogden, 2010) where the smoothness of the data is considered. Random field methods are established as the core inferential tool in neuroimaging studies. These methods correct the statistical thresholds based on the smoothness estimated from the the images, resulting in increased sensitivity. Functional based methods explicitly use smooth fields parametrized by smooth basis functions in the model, thereby regularizing the solution and simplifying the estimation at the same time. Related to functional methods are tensor-based methods (Zhou et al., 2013; Li, 2014) and wavelet-based methods (Van De Ville et al., 2007; Wang et al., 2014), where either low rank tensor factorization or a wavelet basis is used to approximate the spatial field of interest.

Long overlooked in neuroimaging studies, including imaging-genetics, is the influence from unobservable latent factors (Bhattacharya et al., 2011; Montagna et al., 2012). An illustrative cartoon is presented in Figure 1 for a typical neuroimaging-genetic case, in which the effect of interest

is usually small compared with the total variance. This is known as low *signal to noise ratio* (SNR). Large-scale multi-center collaborations have become a common practice in the neuroimaging community (Jack et al., 2008; Consortium et al., 2012; Van Essen et al., 2013; Thompson et al., 2014) and increasing numbers of researchers are starting to pool data from different sources. The heterogeneity of the data introduces large unexplained variance originating from population stratification or cryptic relatedness, for example genetic background, medical history, traumatic experiences and environmental impacts. Such variance aggregates the SNR issue and confuses the estimation procedures if unaccounted for. However these confounding factors are usually difficult or costly to quantify, and therefore they are hidden from the data analysis in most, if not all, studies.

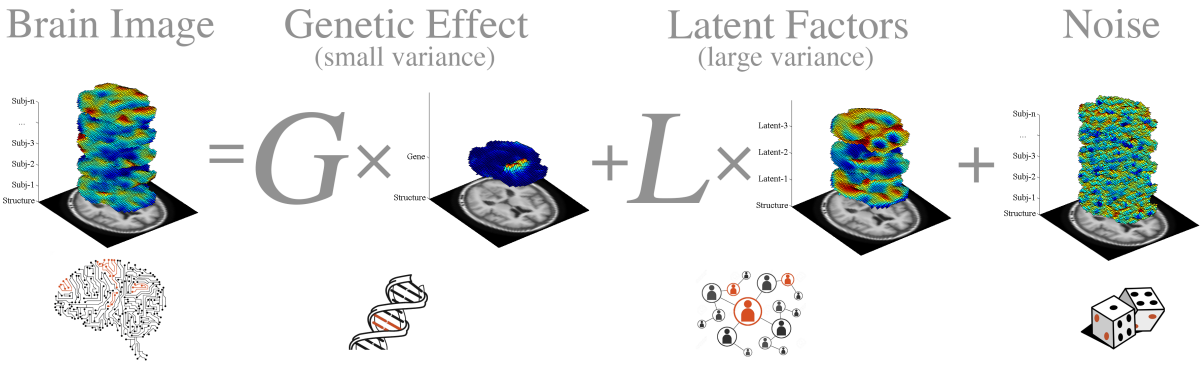


Figure 1: An illustrative cartoon for latent influence in imaging-genetic studies. Low variance genetic effects could be dominated by large variance latent effects. (For simplicity we omit the fixed effect term from the covariates in this illustrative cartoon.)

To see how the latent factor-induced variance undermines the power of statistical procedures, let us take the most commonly used least squares regression as an example. Assume the model $Y = X\beta + L + E$, where Y is the response, X is the predictor of interest, β the regression coefficient, L is the unobservable latent factor and E is the noise term. In the absence of knowledge regarding L, the alternative model $Y = X\widetilde{\beta} + \widetilde{E}$ is estimated instead, where $\widetilde{E} = L + E$. Assuming independence between X, L and E, we have $\text{var}[\widetilde{E}] = \text{var}[L] + \text{var}[E]$, where $\text{var}[\cdot]$ measures the variance. Denote $\widehat{\beta}$ the oracle estimator where the true model is fit with the knowledge of L and $\widehat{\widetilde{\beta}}$ the estimator for the alternative model, the asymptotic theory of least square estimators tells us $\widehat{\beta} \sim$

$\mathcal{N}\left(\beta, \text{var}[E]\,(X'X)^{-1}\right)$ and $\widehat{\widetilde{\beta}} \sim \mathcal{N}\left(\beta, \text{var}[\widetilde{E}]\,(X'X)^{-1}\right)$ as the sample size goes to infinity, that is to say $\widehat{\widetilde{\beta}}$ is more variable than $\widehat{\beta}$ and converges slowly to the population mean. See Figure 2 for a graphical illustration.

Solutions have been proposed to alleviate the loss of statistical efficiency caused by latent factors. In Zhu et al. (2014) the authors propose to dynamically estimate the latent factors from the observed data. However this approach is based on *Markov chain Monte-Carlo* (MCMC) sampling, and therefore the computational cost is prohibitive for high dimensional tensor field applications. In the eQTL literature, several methods that explicitly account for the hidden determinants have been developed. Following a Bayesian formulation, Stegle et al. (2010) integrates out the hidden effect; Fusi et al. (2012) however, computes the ML estimate of hidden factors by marginalizing out the regression coefficients and then using the estimated hidden factors to construct certain covariance matrices for subsequent analyses. These studies are not concerned with the spatial structure and the inherent dimensionality of the model, and the results depend on the choice of parameters for the prior distributions. Additionally, these studies consider latent effect as "variance of no interest", but as we will see in later sections, the latent structure also contains vital information and therefore should not be simply disregarded as unwanted variance.

In this article, we formulate a new generalized reduced rank latent factor regression model (GRRLF) for high dimensional tensor fields. Our method exploits the spatial structure of the neuroimaging data and the low rank structure of the regression coefficient matrix, which computes the effective covariate space, improves the generalization performance and leads to efficient estimation. The model works for general tensor field responses which include a wide range of imaging modalities, *i.e.* MRI, EEG, PET, etc. Although motivated by imaging-genetic applications, the proposed GRRLF is thus widely applicable to almost all types of neuroimaging studies. The estimation is carried out via minimizing a properly defined loss function, which includes *maximum likelihood estimation* (MLE) and *penalized likelihood estimation* (PLE) as special cases.

The contributions of this paper are four-fold. Firstly, we introduce field-constrained latent factor estimation for high dimensional tensor field regression analysis. It efficiently explains the

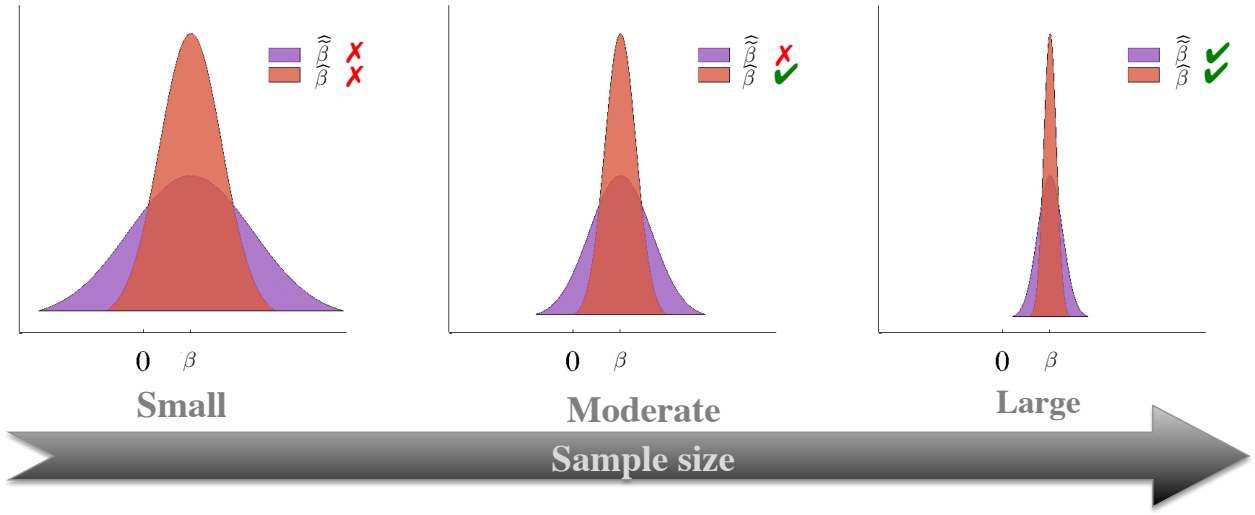


Figure 2: An illustrative example for how the latent factor induced variance undermines the statistical efficiency of least squares estimator. The color coded region are the distribution of the oracle estimator $\widehat{\beta}$ (red) and the alternative estimator $\widehat{\widehat{\beta}}$ (purple) under different sample sizes, with the nonzero population mean β. The oracle estimator requires smaller sample size to achieve the desired sensitivity.

covariance structure in the data caused by the hidden structures. Secondly, our model integrates dimension reduction, that not only improves the statistical efficiency but also facilitates model interpretability. Thirdly, we provide several implementations to efficiently compute the solution under constraints, including *Riemannian manifold optimization* (Absil et al., 2009) and *nuclear norm regularization* which are both based on manifold optimization. We highlight the flexibility of using manifold optimization to formulate neuroimaging problems, which can lead to further interesting applications. Lastly, we present an efficient kernel approach for brain-wide genome-wide association studies under the GRRLF framework and apply it to the ADNI dataset. Empirical results provide evidence that the kernel GRRLF approach is capable of capturing the interactions that can be missed in conventional studies.

The rest of the paper is organized as follows. In the Materials and methods section, we detail the model formulation and estimation. In the Results section, the proposed method is evaluated with both synthetic and real-world examples and compared with other conventional approaches. Finally

we conclude this paper with a summary and future prospects in the Discussion section. The real-world data used and detailed preprocessing steps are described in the Appendix. MATLAB scripts for GRRLF are available online from http://github.com/chenyang-tao/grrlf/.

2 Materials and methods

2.1 Model formulation

Denote the Ω as the spatial domain of the brain and $\boldsymbol{v}$ as its spatial index, $\mathcal{X}, \mathcal{Y}$ are the random vectors/fields of covariates and responses, we denote $\boldsymbol{X} = \{\boldsymbol{x}_i\}_{i=1}^n$ and $\boldsymbol{Y} = \{\boldsymbol{y}_{i,\boldsymbol{v}} | i = 1, \cdots, n, \boldsymbol{v} \in \Omega\}$ the respective empirical sample where $\boldsymbol{x} \in \mathbb{R}^p$, $\boldsymbol{y}_{i,\boldsymbol{v}} \in \mathbb{R}^q$ and n is the sample size. Here p is the dimension of covariates and q is the number of image modalities (for example, $\boldsymbol{y}_{i,\boldsymbol{v}}$ is the 3×3 diffusion tensor from DTI imaging, the 3×1 tissue composition (WM, GM, CSF) from VBM analysis or the time series of a task response). All $\boldsymbol{B} \in \mathbb{R}^{p \times d}$ orthonormal matrices, *i.e.* $\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{I}_d$, form a Riemannian manifold known as the *Stiefel manifold* and denoted as $\mathcal{S}_d(\mathbb{R}^p)$ while a less restrictive manifold requiring only $\mathrm{diag}(\boldsymbol{B}^\top \boldsymbol{B}) = \boldsymbol{I}_d$ is called the *oblique manifold* with the notation $\mathcal{O}_d(\mathbb{R}^p)$. We call d the effective dimension of $\mathcal{X}$ *w.r.t* to $\mathcal{Y}$ if $\mathcal{X} \perp\!\!\!\perp \mathcal{Y} | B^\top \mathcal{X}$ for some projection matrix $B \in \mathcal{S}_d(\mathbb{R}^p)$ where $\perp\!\!\!\perp$ stands for independence and $\cdot|\cdot$ is the conditioning operator. The voxel-wise model writes

$$\boldsymbol{y}_{i,\boldsymbol{v}} = \widetilde{\boldsymbol{\Phi}}_{\boldsymbol{v}} \boldsymbol{B}^\top \boldsymbol{x}_i + \widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{v}} \boldsymbol{l}_i + \boldsymbol{\xi}_{i,\boldsymbol{v}}, \tag{1}$$

where $\boldsymbol{x}$ is the covariate term, $\boldsymbol{l} \in \mathbb{R}^t$ is the latent factor, $\boldsymbol{\xi}_{\boldsymbol{v}} \in \mathbb{R}^q$ is the noise, $\widetilde{\boldsymbol{\Phi}}_{\boldsymbol{v}} \in \mathbb{R}^{q \times d}$ is the covariate regression coefficient and $\widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{v}} \in \mathbb{R}^{q \times t}$ the latent factor loading matrix.

To understand model (1), let us consider a concrete example. Say for example, a researcher is interested in how substance abuse alters brain morphometry. The researcher has collected voxel-wise gray matter and white matter volumes (response $\boldsymbol{y}_{\boldsymbol{v}} \in \mathbb{R}^2$), and various evaluation scores related to substance abuse, including the *Alcohol Use Disorders Identification Test* (AUDIT) (Saun-

ders et al., 1993), *Fagerstrom Test for Nicotine Dependence* (FTND) (Heatherton, 1991) and *Substance Use Risk Profile Scale* (SURPS) (Woicik et al., 2009) for a group of subjects. Each of these evaluations has several sub-scores and altogether the researcher has a 14 dimensional feature vector for each subject (covariate $\boldsymbol{x} \in \mathbb{R}^{14}$). These features are correlated, and it is expected that a low dimensional summary (effective covariate $\widetilde{\boldsymbol{x}} = \boldsymbol{B}^\top \boldsymbol{x} \in \mathbb{R}^d, d \in [1, \cdots, 3]$) is sufficient to explain the variations in brain morphometry caused by substance abuse. The researcher also collects covariates of no interest, such as age, gender and race, that correlate with the imaging features and will be modeled to remove their effect. The researcher is aware that population stratification and subjects' medical history can affect brain tissue volumes, but unfortunately, the subjects are not genotyped and their individual files do not cover medical records therefore such information is unavailable (latent status $\boldsymbol{l}$).

For notational simplicity hereafter we assume $q = 1$ so that we can write the brain-wide model in matrix form. Denote N_{vox} the number of voxels within Ω, then with $\boldsymbol{Y} \in \mathbb{R}^{n \times N_{\text{vox}}}$ the observation matrix, $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ the covariate matrix, $\widetilde{\boldsymbol{\Phi}} \in \mathbb{R}^{p \times N_{\text{vox}}}$ the covariate effect, $\boldsymbol{L} \in \mathbb{R}^{n \times t}$ the latent status matrix, $\widetilde{\boldsymbol{\Gamma}} \in \mathbb{R}^{t \times N_{\text{vox}}}$ the latent response and $\boldsymbol{E} \in \mathbb{R}^{n \times N_{\text{vox}}}$ the noise term, we have the matrix form of the brain-wide model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}\widetilde{\boldsymbol{\Phi}} + \boldsymbol{L}\widetilde{\boldsymbol{\Gamma}} + \boldsymbol{E}, \tag{2}$$

In the case $\{\boldsymbol{\xi}_{\boldsymbol{v}}\}$ are *i.i.d* Gaussian variables, the maximal likelihood solution of GRRLF is

$$\begin{aligned} \left\{\widehat{\widetilde{\boldsymbol{\Phi}}}, \widehat{\boldsymbol{B}}, \widehat{\widetilde{\boldsymbol{\Gamma}}}, \widehat{\boldsymbol{L}}\right\} &= \underset{B, \widetilde{\Phi}, \widetilde{\Gamma}, L}{\arg\min} \left\| \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}\widetilde{\boldsymbol{\Phi}} - \boldsymbol{L}\widetilde{\boldsymbol{\Gamma}} \right\|_{\text{F}}^2 \\ &\text{subject to } \boldsymbol{B} \in \mathcal{S}_d(\mathbb{R}^p) \text{ and } \boldsymbol{L} \in \mathcal{O}_t(\mathbb{R}^n), \end{aligned} \tag{3}$$

where $\|\cdot\|_{\text{Fro}}$ is the Frobenius norm. We note that the restriction on $\boldsymbol{L}$ is simply a normalization and $(\boldsymbol{B}, \widetilde{\boldsymbol{\Phi}})$ is an equivalent class under orthogonal transformations, *i.e.* if $(\boldsymbol{B}, \boldsymbol{\Phi})$ is a solution then $(\boldsymbol{B}\boldsymbol{Q}, \boldsymbol{Q}^\top\boldsymbol{\Phi})$ is also a solution for all unitary matrices $\boldsymbol{Q} \in \mathbb{R}^{d \times d}$. More generally, GRRLF can be formulated as

$$\left\{\widehat{\widetilde{\boldsymbol{\Phi}}}, \widehat{\boldsymbol{B}}, \widehat{\widetilde{\boldsymbol{\Gamma}}}, \widehat{\boldsymbol{L}}\right\} = \underset{B,\widetilde{\Phi},\widetilde{\Gamma},L}{\arg\min}\, \ell(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{B}, \widetilde{\boldsymbol{\Phi}}, \boldsymbol{L}, \widetilde{\boldsymbol{\Gamma}}) \quad \text{subject to } \boldsymbol{B} \in \mathcal{M}_1, \boldsymbol{L} \in \mathcal{M}_2, \tag{4}$$

where ℓ is some loss function and $\{\mathcal{M}_i\}_{i=1}^2$ are some Riemannian manifolds to constrain the solution.

2.2 Smoothing the tensor fields

To more effectively exploit the spatial structures, further constraints can be enforced. For example, it is natural to assume the smoothness of tensor fields $\widetilde{\boldsymbol{\Phi}}$ and $\widetilde{\boldsymbol{\Gamma}}$. In this work, we assume $\widetilde{\boldsymbol{\Phi}}$ and $\widetilde{\boldsymbol{\Gamma}}$ can be approximated by linear combinations of some (smooth) basis functions as

$$\widetilde{\boldsymbol{\Phi}}_{\boldsymbol{v}} = \sum_{b=1}^{N_{\text{knot}}} h_b(\boldsymbol{v}) \boldsymbol{\Phi}_b, \; \widetilde{\boldsymbol{\Gamma}}_{\boldsymbol{v}} = \sum_{b=1}^{N_{\text{knot}}} h_b(\boldsymbol{v}) \boldsymbol{\Gamma}_b,$$

where $\{h_b(\cdot)\}_{b=1}^{N_{\text{knot}}}$ is the set of basis functions, $\{\boldsymbol{\Phi}_b \in \mathbb{R}^{q\times d}\}$ and $\{\boldsymbol{\Gamma}_b \in \mathbb{R}^{q\times t}\}$ are the coefficients, and here we have assumed both tensor fields have the same "smoothness" for notational clarity. Similarly to model (2) the smoothed model can be written in matrix form as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}\boldsymbol{\Phi}\boldsymbol{H} + \boldsymbol{L}\boldsymbol{\Gamma}\boldsymbol{H} + \boldsymbol{E}, \tag{5}$$

where $\boldsymbol{\Phi} \in \mathbb{R}^{d\times N_{\text{knot}}}$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{t\times N_{\text{knot}}}$ are the coefficient matrices, and we call $\boldsymbol{H} \in \mathbb{R}^{N_{\text{knot}}\times N_{\text{vox}}}$ the smoothing matrix. $\widetilde{\boldsymbol{\Phi}} = \boldsymbol{\Phi}\boldsymbol{H}$ and $\widetilde{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\boldsymbol{H}$ are respectively referred to as the covariate response field and the latent response field, $\widetilde{\boldsymbol{B}} = \boldsymbol{B}\boldsymbol{\Phi} \in \mathbb{R}^{p\times N_{\text{knot}}}$ as the covariate effect matrix and $\widetilde{\boldsymbol{L}} = \boldsymbol{L}\boldsymbol{\Gamma} \in \mathbb{R}^{n\times N_{\text{knot}}}$ as the latent effect matrix. Since $N_{\text{knot}} \ll N_{\text{vox}}$, the smoothing operation can significantly reduce the number of parameters to be optimized. In this study we have used Gaussian radial basis functions (RBF) $\{h_b(\boldsymbol{v}) = \exp(-\|\boldsymbol{v} - \boldsymbol{v}_b\|_2^2/2\sigma^2)\}_{b=1}^{N_{\text{knot}}}$ as basis functions, where $\boldsymbol{v}_b \in \Omega$ is the b-th knot and σ^2 is the bandwidth parameter. Other non-smooth basis functions can also be used if they

are well justified for the application.

Note that (5) is a very general formulation that encompasses many statistical models as special cases. In Table S1 we provide an inexhaustive list of loss function ℓ that lead to commonly used statistical models.

2.3 Generalized cross-validation procedure

GRRLF needs constraint parameters $\boldsymbol{\Theta}$ to regularize its solution, therefore a parameter selection procedure is necessary to ensure good generalization performance. In the literature, a *cross-validation* (CV) procedure is often used to assess the generalizing performance of the parameters, by evaluating the loss with the validation set and the parameters estimated from the training set. However, for GRRLF the conventional CV procedure can not be used, because the latent parameters are unique to the validation set, and as such, they can not be estimated from the training set. Here we propose a *generalized cross-validation* (GCV) procedure to resolve this dilemma.

Assuming the training and validation sets are drawn from the same distribution, we know that for the latent component the latent response field $\widetilde{\boldsymbol{\Gamma}}$ is shared by both sets while the latent status variables $\boldsymbol{L}$ are different. Therefore given $\{\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{\Gamma}}\}$, we can estimate the latent status $\boldsymbol{L}$ of the validation set by minimizing the residual error $\left\|\boldsymbol{Y}_{\text{test}} - \boldsymbol{X}_{\text{test}}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{\Phi}}\boldsymbol{H} - \boldsymbol{L}\widehat{\boldsymbol{\Gamma}}\boldsymbol{H}\right\|_F^2$ of the validation set and using the minimal residual error as the generalizing performance score. The pseudo code for GCV is given in Algorithm 1.

2.4 Estimation based on Riemannian manifold optimization

Since (5) is a nonlinear optimization problem constrained on Riemannian manifolds it is difficult to optimize directly. A key observation is that individually optimizing $\{\boldsymbol{B}, \boldsymbol{\Phi}, \boldsymbol{L}, \boldsymbol{\Gamma}\}$ reduces to a linear problem, which suggests the use of the so-called *block relaxation algorithm* (De Leeuw, 1994; Lange, 2010) to alternately update $\{\boldsymbol{B}, \boldsymbol{\Phi}, \boldsymbol{L}, \boldsymbol{\Gamma}\}$ at each iteration until convergence. In this work, we use the *manopt* toolbox (Boumal et al., 2014) to efficiently solve the manifold optimization

Algorithm 1: Generalized cross-validation procedure for GRRLF

Input: $\boldsymbol{X}_{\text{eval}}, \boldsymbol{Y}_{\text{eval}}, \boldsymbol{X}_{\text{test}}, \boldsymbol{Y}_{\text{test}}, \boldsymbol{H}, \boldsymbol{\Theta}$

Output: Err_{CV}

$\left(\widehat{\widetilde{\boldsymbol{B}}}, \widehat{\widetilde{\boldsymbol{L}}}\right) := \text{GRRLF}\left(\boldsymbol{X}_{\text{eval}}, \boldsymbol{Y}_{\text{eval}}, \boldsymbol{H}, \boldsymbol{\Theta}\right)$;

$[\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V}] := \text{SVD}\left(\widehat{\widetilde{\boldsymbol{L}}}\right), \widehat{\boldsymbol{\Gamma}} := \boldsymbol{V}^{\top}$; /* $\widehat{\widetilde{\boldsymbol{L}}} := \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top}$ */

$\widehat{\widetilde{\boldsymbol{R}}}_{\text{test}} := \boldsymbol{Y}_{\text{test}} - \boldsymbol{X}_{\text{test}}\widehat{\widetilde{\boldsymbol{B}}}_{\text{eval}}\boldsymbol{H}$;

$\widehat{\boldsymbol{L}}_{\text{test}} := \arg\min_{L} \left\| \widehat{\widetilde{\boldsymbol{R}}}_{\text{test}} - \boldsymbol{L}\widehat{\boldsymbol{\Gamma}}\boldsymbol{H} \right\|_F^2$; /* Least squares */

$\text{Err}_{\text{CV}} := \left\| \widehat{\widetilde{\boldsymbol{R}}}_{\text{test}} - \widehat{\boldsymbol{L}}_{\text{test}}\widehat{\boldsymbol{\Gamma}}\boldsymbol{H} \right\|_F^2$;

problem (5). Manopt provides a general framework for solving Riemannian manifold optimization problems, which grants the modeler the freedom of specifying the constraints for the model without worrying about the implementation details and still use an efficient solver. We remark that while the general purpose solver relieves the burden from the modeler, the computational efficiency can be significantly improved using a customized solver *w.r.t* the loss ℓ. Here we detail the implementation details of our customized solver for (4).

The key idea of GRRLF is to improve the estimation of weak signals by accounting for the strong signals. Therefore, if the covariate signal or the latent signal is of interest, and there is no prior knowledge of which signal is "dominating", a choice on which component is used to start the iteration should be made. For example, if the covariate effect is dominating but the latent effect is estimated first, then part of the covariate effect might be erroneously interpreted as latent effect. Here we propose to base our decision on the generalizing performance from GCV. If the 'latent first' strategy is favored by GCV, we further test for the association between covariates and estimated latent status variables, using dependency tests such as CCA or more general Hilbert-Schmidt independence criteria (HSIC) (Gretton et al., 2007). If a significant association is detected between the covariates and the latent status variables, the previous decision is overruled and, instead, we estimate the covariate effect first. The complete estimation procedure is summarized in Algorithm 2, and hereafter we refer to it as the general manifold GRRLF implementation (GM-GRRLF). More sophisticated procedures, that control for the dependency between covariates

Algorithm 2: GM-GRRLF

Input: $\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{H}, \mathcal{M}_1, \mathcal{M}_2$
Output: $\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Phi}}, \widehat{\boldsymbol{L}}, \widehat{\boldsymbol{\Gamma}}$
Initialize: $\widehat{\boldsymbol{B}}_0, \widehat{\boldsymbol{\Phi}}_0, \widehat{\boldsymbol{L}}_0, \widehat{\boldsymbol{\Gamma}}_0$
`/* Decide which component to estimate first */`
if *Covariate first* **then**
  $[\widehat{\boldsymbol{B}}_0, \widehat{\boldsymbol{\Phi}}_0] :=$ **SolveCovariate**$(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{H}, \widehat{\boldsymbol{B}}_0, \widehat{\boldsymbol{\Phi}}_0, \mathcal{M}_2)$;
end if
while *Stopping criteria are not met* **do**
  $[\widehat{\boldsymbol{L}}_i, \widehat{\boldsymbol{\Gamma}}_i] :=$ ***SolveLatent***$(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}_{i-1}\widehat{\boldsymbol{\Phi}}_{i-1}\boldsymbol{H}, \boldsymbol{H}, \widehat{\boldsymbol{L}}_{i-1}, \widehat{\boldsymbol{\Gamma}}_{i-1}, \mathcal{M}_1)$;
  $[\widehat{\boldsymbol{B}}_i, \widehat{\boldsymbol{\Phi}}_i] :=$ ***SolveCovariate***$(\boldsymbol{Y} - \widehat{\boldsymbol{L}}_i\widehat{\boldsymbol{\Gamma}}_i\boldsymbol{H}, \boldsymbol{X}, \boldsymbol{H})$;
end while
$\widehat{\boldsymbol{B}} := \widehat{\boldsymbol{B}}_{\text{last}}, \widehat{\boldsymbol{\Phi}} := \widehat{\boldsymbol{\Phi}}_{\text{last}}, \widehat{\boldsymbol{L}} := \widehat{\boldsymbol{L}}_{\text{last}}, \widehat{\boldsymbol{\Gamma}} := \widehat{\boldsymbol{\Gamma}}_{\text{last}}$;

Function $[\widehat{\boldsymbol{B}}, \widehat{\boldsymbol{\Phi}}] :=$ ***SolveCovariate***$(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{H}, \boldsymbol{B}_0, \boldsymbol{\Phi}_0, \mathcal{M})$**:**
  $\widehat{\boldsymbol{B}}_0 = \boldsymbol{B}_0, \widehat{\boldsymbol{\Phi}}_0 = \boldsymbol{\Phi}_0, i := 0$;
  while *Stopping criteria are not met* **do**
    $i := i + 1$;
    $\widehat{\boldsymbol{B}} := \arg\min_{B \in \mathcal{M}} \left\|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}\widehat{\boldsymbol{\Phi}}_0\boldsymbol{H}\right\|_F^2$; `/* Manifold optimization */`
    $\widehat{\boldsymbol{\Phi}} := \arg\min_{\Phi} \left\|\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}_0\boldsymbol{\Phi}\boldsymbol{H}\right\|_F^2$; `/* Least squares */`
  end while
end
Function $[\widehat{\boldsymbol{L}}, \widehat{\boldsymbol{\Gamma}}] =$ ***SolveLatent***$(\boldsymbol{Y}, \boldsymbol{H}, \boldsymbol{L}_0, \boldsymbol{\Gamma}_0, \mathcal{M})$**:**
  $\widehat{\boldsymbol{L}}_0 := \boldsymbol{L}_0, \widehat{\boldsymbol{\Gamma}}_0 := \boldsymbol{\Gamma}_0, i = 0$;
  while *Stopping criteria are not met* **do**
    $i := i + 1$;
    $\widehat{\boldsymbol{\Gamma}} := \arg\min_{\Gamma \in \mathcal{M}} \left\|\boldsymbol{Y} - \widehat{\boldsymbol{L}}_{i-1}\boldsymbol{\Gamma}\boldsymbol{H}\right\|_F^2$; `/* Manifold optimization */`
    $\widehat{\boldsymbol{L}} := \arg\min_{L} \left\|\boldsymbol{Y} - \boldsymbol{L}\widehat{\boldsymbol{\Gamma}}_{i-1}\boldsymbol{H}\right\|_F^2$; `/* Least squares */`
  end while
end

and latent components, are discussed in later sections.

2.5 Model selection

The performance of GRRLF depends on the parameters (d, t), denoting the effective dimension of the covariates and latent factors. In the absence of prior knowledge of (d, t), we can use the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz

et al., 1978) or generalized cross-validation described above to dynamically determine these two parameters. Likelihood models can use *IC (and possibly also a χ^2-test) to select (d, t) while for other more general models the GCV approach is preferred. For (4), fast determination of (d, t) can be achieved by combining RRR and PCA. The sequence of RRR and PCA is determined based on generalized cross validation. Both RRR and PCA involve solving an eigenvalue problem and the magnitude of the eigenvalues provides information about the inherent structural dimensionality of the data. Assuming the eigenvalues estimated are sorted in descending order, the 'elbow' or 'jump point' of the eigenvalue curve is used as an estimate of the rank/structural dimensionality.

2.6 Constrained nuclear norm formulation of GRRLF

In this section we present an alternative formulation of GRRLF using the nuclear norm regularization (NNR), which has a global optima and can be solved with convex optimization techniques. NNR is a powerful tool restoring the low rank structure of matrices with noisy or incomplete observations and is widely used in machine learning applications (Yuan et al., 2007; Koren et al., 2009; Candès and Tao, 2010).

Notice that solving (5) is a nonlinear optimization problem, and its solutions can be easily trapped in local optima. However, solving model

$$\boldsymbol{Y} = \boldsymbol{X}\widetilde{\boldsymbol{B}}\boldsymbol{H} + \widetilde{\boldsymbol{L}}\boldsymbol{H} + \boldsymbol{E} \tag{6}$$

for convex loss function $f(\cdot)$ with respect to $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{L}}$ is easy because there exists a global minimum that can be easily approached with standard optimization tools. But this nice property is no longer valid with the rank constraints applied, because: 1) the manifold has changed and the geodesics is different, so $f(\cdot)$ may no longer be convex; 2) the feasible solution domain may no longer be a convex set. To overcome such difficulties, an alternative formulation that produces effective low rank solutions while keeping the convexity of the problem is desired. NNR fulfills such needs.

For a matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$, its *nuclear norm* (NN) $\|\boldsymbol{A}\|_*$ is defined as the ℓ_1 norm of its singular values $\{\sigma_h\}_{h=1}^{\min(n,m)}$, or equivalently as $\|\boldsymbol{A}\|_* = \mathrm{tr}(\sqrt{\boldsymbol{A}\boldsymbol{A}^\top})$ thus also known as the *trace norm*. It can be shown for matrix completion problems, penalizing the nuclear norm of the solution matrices is equivalent to thresholding the singular values thus producing low rank results (Chen et al., 2013). Thus for GRRLF, we can similarly write down its NNR formulation as

$$(\widehat{\widetilde{\boldsymbol{B}}}, \widehat{\widetilde{\boldsymbol{L}}}) = \underset{\widetilde{\boldsymbol{B}},\widetilde{\boldsymbol{L}}}{\arg\min} \left\|\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{B}}\boldsymbol{H} - \widetilde{\boldsymbol{L}}\boldsymbol{H}\right\|_F^2 + \lambda_1 \left\|\widetilde{\boldsymbol{B}}\right\|_* + \lambda_2 \left\|\widetilde{\boldsymbol{L}}\right\|_*, \tag{7}$$

where λ_1 and λ_2 are regularization parameters. Since $\|\cdot\|_*$ does not admit an analytical expression, in this study we optimize the following alternative form

$$(\widehat{\widetilde{\boldsymbol{B}}}, \widehat{\widetilde{\boldsymbol{L}}}) = \underset{\|\widetilde{\boldsymbol{B}}\|\le t_1, \|\widetilde{\boldsymbol{L}}\|\le t_2}{\arg\min} \left\|\boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{B}}\boldsymbol{H} - \widetilde{\boldsymbol{L}}\boldsymbol{H}\right\|_F^2, \tag{8}$$

where t_1 and t_2 are the NN constraints, therefore (8) becomes a constrained optimization problem. By extending the results of M. Jaggi (2010) we prove that (8) is equivalent to a convex optimization problem on the domain of fixed-trace *positive semi-definite* (PSD) matrices and present the pseudo code for estimation in Algorithm 3. The details are provided in Appendix C.

Algorithm 3: NNR-GRRLF

Input: $\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{H}, t_1, t_2$
Output: $\widehat{\widetilde{\boldsymbol{B}}}, \widehat{\widetilde{\boldsymbol{L}}}$
Set $k := 1$
Initialize $\boldsymbol{Z}_B^{(0)} \in \mathbb{S}_{\text{PSD}}^{p+m}(1)$, $\boldsymbol{Z}_L^{(0)} \in \mathbb{S}_{\text{PSD}}^{n+m}(1)$
while *the stopping criteria are not satisfied* **do**
  Compute $\boldsymbol{v}_B^{(k)} = \mathrm{MaxEV}(-\nabla_{\boldsymbol{B}_t} f_t)$, $\boldsymbol{v}_L^{(k)} = \mathrm{MaxEV}(-\nabla_{\boldsymbol{L}_t} f_t)$
  Compute the optimal learning rate α_k
  Update $\boldsymbol{Z}_B^{(k+1)} := \boldsymbol{Z}_B^{(k)} + \alpha_k(\boldsymbol{v}_B^{(k)}\boldsymbol{v}_B^{(k)\top} - \boldsymbol{Z}_B^{(k)})$
  Update $\boldsymbol{Z}_L^{(k+1)} := \boldsymbol{Z}_L^{(k)} + \alpha_k(\boldsymbol{v}_L^{(k)}\boldsymbol{v}_L^{(k)\top} - \boldsymbol{Z}_L^{(k)})$
  Set $k := k + 1$, correct the solution if necessary
end while
$\widehat{\widetilde{\boldsymbol{B}}} := \left[\boldsymbol{Z}_B^{(k)}\right]_{1:p}^{p+1:p+m}$, $\widehat{\widetilde{\boldsymbol{L}}} := \left[\boldsymbol{Z}_L^{(k)}\right]_{1:n}^{n+1:n+m}$

2.7 An efficient kernel GWAS extension

The model developed so far focuses on the candidate approach, *i.e.* a set of variables of interest are grouped into the candidate predictor $\boldsymbol{x}$ and then we proceed with the model estimation. However, in modern neuroimaging-genetic studies, a *genome wide association study* (GWAS) is often performed, which means testing the association with the imaging phenotypes for a colossal number of candidate genes / SNPs (typically anywhere from thousands to millions). Estimating the complete model for each candidates incurs a heavy computational burden, a practice often to be avoided even in conventional univariate GWAS studies (Eu-ahsunthornwattana et al., 2014). Also an accurate yet efficient statistical testing procedure is required to assign the significance level to the observed association.

Inspired by the works of Liu et al. (2007) and Ge et al. (2012), we propose to address the challenge of an efficient GRRLF-GWAS implementation by integrating the powerful least squares kernel machines (LSKM) under the *small effect assumption*. For convenience we will refer to the genetic candidates as 'genes' in the following exposition. Specifically, the model writes

$$\boldsymbol{y}_{i,\boldsymbol{v}} = h_{\boldsymbol{v}}^{(k)}(\boldsymbol{g}_i^{(k)}) + \boldsymbol{\Phi}_{\boldsymbol{v}}\boldsymbol{B}^{\top}\boldsymbol{x}_i + \Gamma_{\boldsymbol{v}}\boldsymbol{l}_i + \boldsymbol{\xi}_{i,\boldsymbol{v}}^{(k)},$$

where $k \in \{1, \cdots, p_g\}$ is the index for the genes, $\boldsymbol{g}^{(k)}$ is the data for the k-th gene, $h^{(k)}(\cdot)$ is the nonparametric function defined on the k-th genetic data in some function space $\mathcal{H}_k$, $\boldsymbol{\xi}_{i,\boldsymbol{v}}^{(k)}$ is the gene-specific residual component and the rest follows the previous definitions, except that we have shifted our interest from $\boldsymbol{x}$ to $\boldsymbol{g}$. We will suppress the index i, k and $\boldsymbol{v}$ for clarity whenever the context is clear. The function space $\mathcal{H}$ is determined by the semi-positive definite kernel function $\kappa(\boldsymbol{g}, \boldsymbol{g}')$ defined on the genetic data and we call the matrix $\boldsymbol{K}$ defined by $K_{ij} = \kappa(\boldsymbol{g}_i, \boldsymbol{g}_j)$ the kernel matrix. The small effect assumption basically states that the gene induced variance $\text{var}[h(\boldsymbol{g})]$ is small compared with the gene-specific residual variance $\text{var}[\boldsymbol{\xi}]$ and ignoring it does not significantly bias the estimation for non-genetic parameters in the model. So, instead of estimating the complete model for each gene, the non-genetic parameters are estimated once under the full null

model where $h^{(k)}$ equals zero for all k. Then a score test is performed using the empirical kernel matrix $K^{(k)}$ and the estimated residual component $\widehat{\boldsymbol{\xi}}$ for each gene k, for example, in the case of univariate response,

$$Q^{(k)} := \frac{1}{2\widehat{\sigma}^2}\widehat{\boldsymbol{\xi}}^\top K^{(k)}\widehat{\boldsymbol{\xi}},$$

where $Q^{(k)}$ is the test score following a mixed chi-square distribution under the null hypothesis with some mild conditions and $\widehat{\sigma}^2$ is the estimated variance of the residual $\boldsymbol{\xi}$. The mixed chi-square is approximated by a scaled chi-square with moment matching and the significance level is assigned based on the parametric approximation (Hua and Ghosh, 2014). Note however that the validity of using the parametric approximation hinges on its closeness to the null distribution, which should always be examined in practice. If the approximation deviates from the empirical null, the later should be used. Statistical correction procedures should be invoked after the computation of significance maps to control for the false positives. For example, Bonferroni or FDR can be used for the gene-wise correction, and the peak inference or cluster size inference for the spatial correction. Consult Appendix H for detailed discussions.

2.8 Independence between the covariate effect and the latent effect

In some applications the independence between the covariate effect and the latent effect is assumed. In the simplest case of two zero mean Gaussian variables ξ and ζ, independence is equivalent to vanishing covariance between the variables, *i.e.* $\mathrm{cov}[\xi, \zeta] = 0$. For their empirical samples $\boldsymbol{\xi}, \boldsymbol{\zeta} \in \mathbb{R}^n$ and $\boldsymbol{\zeta}$, this means the asymptotic orthogonality $\lim_{n\to\infty} n^{-1}\boldsymbol{\xi}^\top\boldsymbol{\zeta} = 0$. Now let us assume covariate variable $X \in \mathbb{R}^p$ and latent status $L \in \mathbb{R}^t$ are jointly zero mean Gaussian variables and their covariance matrices are of full rank. Then for their empirical sample $\boldsymbol{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{L} \in \mathbb{R}^{n\times t}$, the orthogonality condition writes $\boldsymbol{X}^\top\boldsymbol{L} = \boldsymbol{O}$ and $\boldsymbol{L}^\top\mathbf{1}_n = \mathbf{0}$, where the columns of $\boldsymbol{X}$ have already been centralized. This brings $(p+1)\times t$ linear equality constraints to $\boldsymbol{L}$ so it can be reparameterized to $\boldsymbol{L}' \in \mathbb{R}^{(n-p-1)\times t}$, then we restrict $\boldsymbol{\Gamma}$ instead of $\boldsymbol{L}'$ to some bounded manifold (for example the

oblique manifold) and carry out the GM-GRRLF estimation.

For more general cases, for example non-Gaussian state variables, we propose to encourage the independence by penalizing the loss function (likelihood in most cases) with a measure of dependency $\Upsilon(\cdot,\cdot)$ between the covariate variable X and latent status L, which generalizes the concept of "orthogonality" in the Gaussian case. More specifically, we optimize the model

$$\ell(\boldsymbol{B},\boldsymbol{\Phi},\boldsymbol{L},\boldsymbol{\Gamma}|\boldsymbol{X},\boldsymbol{Y})+\lambda\Upsilon(\boldsymbol{X},\boldsymbol{L}), \tag{9}$$

where λ is the regularization parameter that balances the trade off. A good candidate for $\Upsilon(\cdot,\cdot)$ is the square loss mutual information (Karasuyama and Sugiyama, 2012). We note however, $\Upsilon(\cdot,\cdot)$ usually has its own parameter to be optimized, and solving (9) can be extremely expensive.

3 Results

3.1 Synthetic examples

For clarity, we use a 1-D synthetic example to illustrate the proposed method[1]. The synthetic data are generated as follows: $N_{\text{knot}}=10$ knots and $N_{\text{vox}}=100$ artificial voxels are placed uniformly on interval $\Omega=[0,1]$ and kernel bandwidth set to $\sigma=0.1$, we set $p=10,q=1,d=2,t=2$, $\boldsymbol{B}=[\boldsymbol{I}_2;\boldsymbol{O}]$ (so only the first two dimensions of the covariate are contributing), $X\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_p)$, $\Phi\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{q\times d\times N_{\text{knot}}}),\Gamma\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_{q\times t\times N_{\text{knot}}}),L\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_t)$ and $\boldsymbol{\xi}_{\boldsymbol{v}}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I}_q)$ independent from other voxels unless otherwise specified. For each simulation $n=100$ samples are drawn. We use nonparametric permutations to obtain the p-values for the sensitivity studies. Specifically, the sum of squared error (sse) is used as the test statistic and the empirical p-value is determined by $p_{\text{emp}}=\max(\#\{\text{sse}_b\le\text{sse}_0\},1)/m_{\text{perm}}$, where $\#\{\cdot\}$ denotes the counting measure, m_{perm} the number of permutation runs, $b=1,\cdots,m_{\text{perm}}$ the permutation index, $\text{sse}_b=\sum_{i,\boldsymbol{v}}\left\|\widehat{e}^b_{i,\boldsymbol{v}}\right\|^2$, $\widehat{e}^b_{i,\boldsymbol{v}}$ denote the residual estimated at voxel $\boldsymbol{v}$ for sample i with the b-th permuted X and $b=0$ refers to the

[1] Imaging a ray shooting through the brain, and we are looking at the responses from the voxels along the trajectory of the ray.

original X.

We first experiment with the NNR implementation of GRRLF. We set the candidate parameter set for nuclear norm constraints t_i to $\{2^0, 2^1, \cdots, 2^{15}\}$, and we stop the iteration when either of the following criteria is satisfied:

1) the number of iterations reaches $k = 3,000$;

2) the improvement of the current iteration is less than 10^{-5} compared with the average of previous 10 iterations.

The performance is evaluated by the *relative mean square error* (RMSE) defined by

$$\text{RMSE} = \frac{\left\|\boldsymbol{A} - \widehat{\boldsymbol{A}}\right\|_F}{\|\boldsymbol{A}\|_F}.$$

Figure 3(a-b) respectively visualizes the optimization procedure and the regularization path of the solution matrices' nuclear norm, and only the results for parameter pairs (t_1, t_2) satisfying $t_1 = t_2$ are shown. For tight constraints (with small t_i), the solutions converge rapidly and the optimal solutions are achieved on the boundary of the feasible domain. Slow convergence is observed for larger t_i, and as the constraints are relaxed the solutions move away from the boundary.

Figure 4 gives an example of the regularization paths of the leading singular values of the NNR-GRRLF solution matrices. To facilitate visualization we have used the normalized SVs defined by $\widetilde{\sigma}_h = \sigma_h / \left(\sum_{h'} \sigma_{h'}\right)$, where $\{\sigma_h\}$ are the original SVs. Under the nuclear norm constraints, the solution matrices show sparsity with respect to their SVs. We call the number of SVs that are bounded away from zero as the "effective rank" (ER) of the matrix; as the nuclear norm constraints are relaxed, ER grows.

Figure 5 gives an example of a GCV RMSE heatmap for parameter selection. The RMSE on the training sample drops as the NN constraints are relaxed, as more flexibilities are allowed for the model. Interestingly for a wide range of parameter settings the RMSE on the validation sample is smaller than that on the training sample, which seems contradictory for CV procedures. This is because with our modified CV procedure,

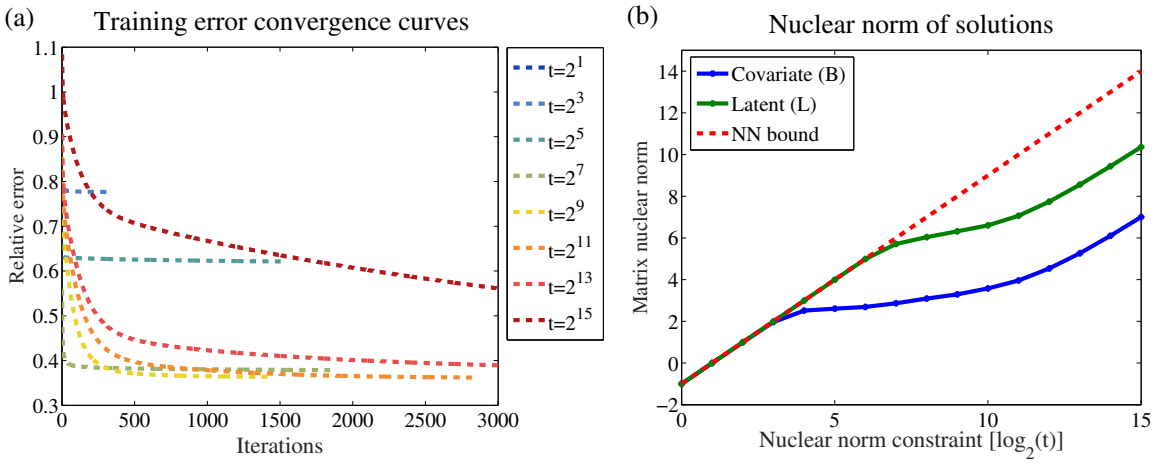


Figure 3: NNR-GRRLF model estimation with Jaggi-Hazan algorithm. (a) Convergence curve of the normalized mean square error for different constraints. (b) Nuclear norm constraint VS nuclear norm of the solution matrices, blue solid line for the covariate coefficient matrix, green solid line for the latent coefficient matrix and red dash line is the nuclear norm upper bound with respect to the constraint. Here we have fixed $t_1 = t_2$.

1) NN of the latent coefficient matrix is no longer bounded;

2) the latent response field $\widetilde{\mathbf{\Gamma}} = \mathbf{\Gamma}\boldsymbol{H}$ is well approximated, although $\widetilde{\boldsymbol{L}} = \boldsymbol{L}\mathbf{\Gamma}$ is not because of the NN constraint.

In practice a relatively large region of the parameter space can show similar good generalization performance (for example see Figure 5(b)). This is because the framework is robust to a small level of over relaxation, and the latent part of the model can compensate for the modeling error from the covariate part, to some extent. In the spirit of Occam's razor, we want to keep the simplest model. This means that the model with the tightest constraints (smallest t_i, with the latent constraint t_2 is prioritized) should be preferred when the validation RMSE is tied.

For GM-GRRLF, we compare AIC, BIC and RRR-PCA for automatic model selection. We perform experiments on the selection of coefficient rank d and latent dimension t. All combinations in $\{(d', t')|d', t' = 1, \cdots, 4\}$ are tested with all experiments repeated for $m = 100$ times and the results are presented in Figure 6. In Figure 6(a), the mean raw score and mean rank of AIC and BIC are shown. AIC gives more confusing results, as it is difficult to choose between $(1, 3)$ and $(2, 2)$. In such ties we opt for the model with the larger coefficient rank because in the absence

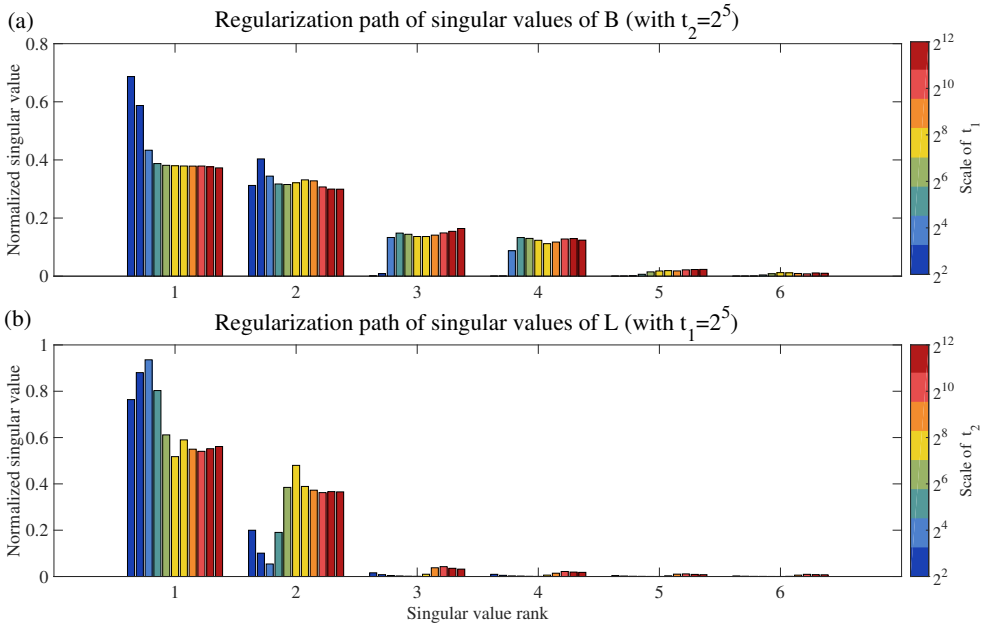


Figure 4: Regularization path for the (normalized) leading singular values with respect to the nuclear norm constraint. (a) Regularization path for t_1 with t_2 fixed. (b) Regularization path for t_2 with t_1 fixed. X-axis spans the six leading singular values and Y-axis indicates their (normalized) magnitude; different regularization parameters are color coded. It can be seen the effective rank of the solution grows with the nuclear norm constraint.

of predictive information, the latent factor part of the model will try to interpret the signal as a latent contribution. AIC also tends to favor models that are larger than the original model. BIC seems to be a better choice as it successfully identifies the true structural dimensionality at its minimum value. As can be seen in Figure 6(b), RRR-PCA also performs well in that it successfully identifies t and narrows down the choice of d to 2 or 3. Taking into account that RRR-PCA is much more computationally efficient than *IC based model selection methods, it is therefore favorable in neuroimaging studies. One can also use the GCV procedure to identify the appropriate model order.

We now compare the two different implementations of GRRLF (GM and NNR) with voxel-wise *least-square regression* (LSR) and whole field *reduced rank regression* (RRR). LSR corresponds to the massive univariate approaches most commonly used in neuroimaging studies, and RRR corresponds to those methods that only consider spatial correlations. For GM-GRRLF and

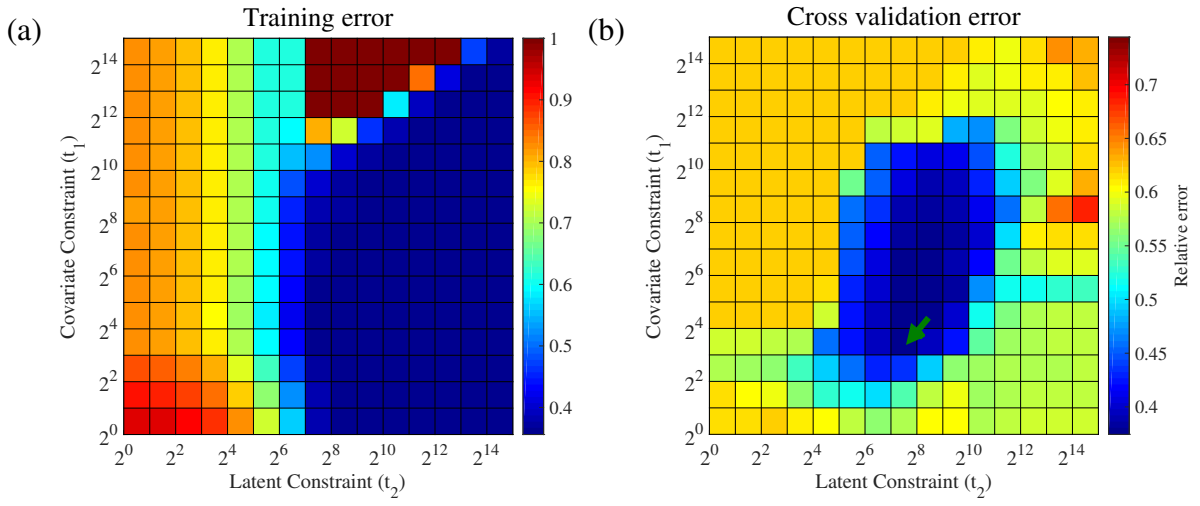


Figure 5: Residual heatmaps for nuclear norm regularization parameter selection. (a) Relative mean square error for the training sample. (b) Relative mean square error for the validation sample. Y-axis corresponds to the covariate coefficient constraint t_1 and X-axis corresponds to the latent coefficient constraint t_2. The green arrow points at the parameter pair with minimal validation error.

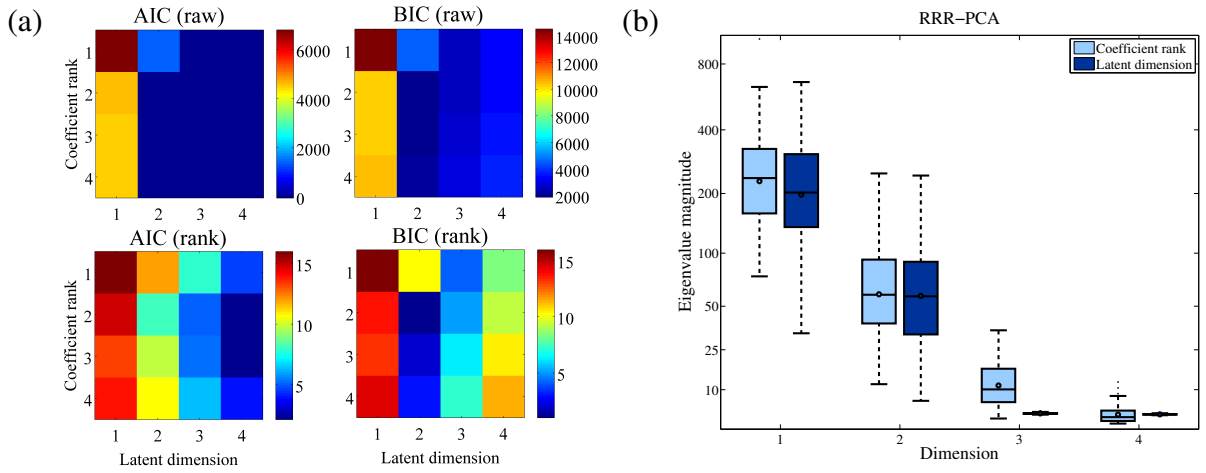


Figure 6: (a) Mean raw and rank map for AIC and BIC. (b) Box plot of eigenvalues from RRR-PCA. *IC procedures identifies model order with lowest score as optimal while RRR-PCA bases its decision on the jumping point of eigenvalues. AIC slightly over estimates the model order and BIC makes the right decision, RRR-PCA gives an fair estimate with much less cost compared with *IC methods.

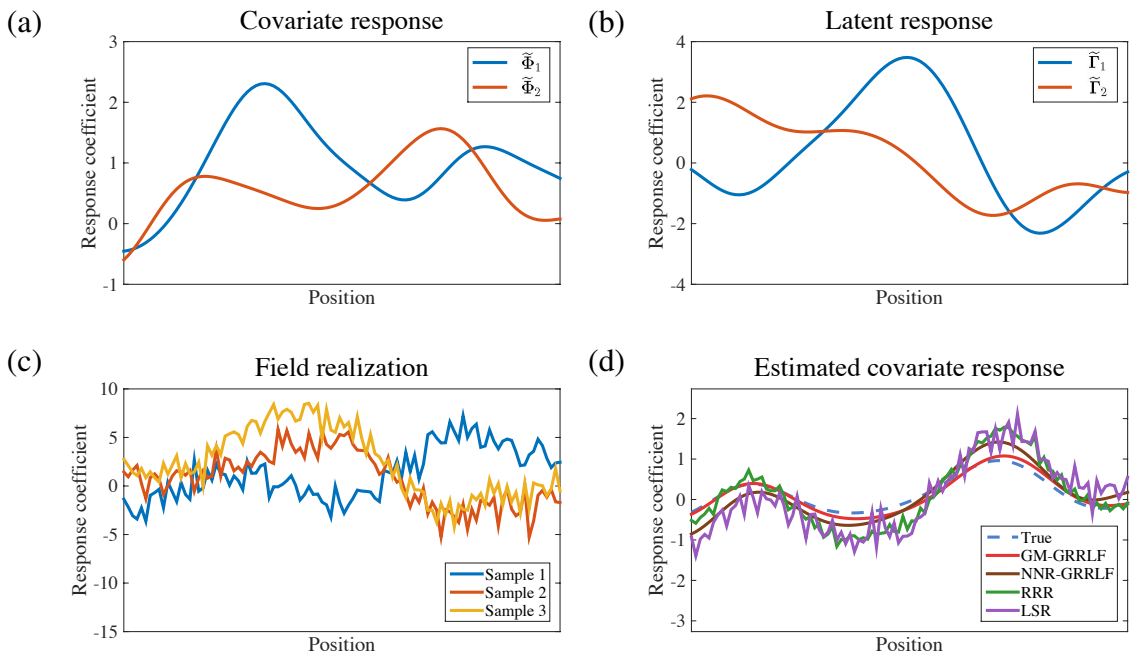


Figure 7: A 1-D example of GRRLF model. (a) True covariate response fields $\widetilde{\mathbf{\Phi}}$. (b) True latent factor response fields $\widetilde{\mathbf{\Gamma}}$. (c) Observed responses for three randomly selected samples. (d) Estimated response using GM-GRRLF, NNR-GRRLF, RRR and LSR together with the ground truth expected response for an unseen sample. (dotted: ground truth, red: GM-GRRLF, brown: NNR-GRRLF, green: RRR, purple: LSR) This demonstrates GRRLF is robust to the latent influences while common solutions fail.

RRR the regression coefficient rank, latent factor number and kernel bandwidth are set to the ground truth. Figure 7 presents an illustrative example: the upper panel gives the smooth response curves corresponding to the effective covariate space and latent space while the lower left figure visualizes three noisy field realizations. In Figure 7(d), the estimated covariate response curves for an unseen sample using different methods are shown. As can be seen, LSR gave the most noisy estimate as it disregards all spatial information while RRR gave a much smoother estimate by considering the covariance structure. However both of them were susceptible to the influence of latent responses, which drove their estimates away from the true response. Overall GRRLF methods showed more robustness against the latent influences, and GM-GRRLF gives the best result. The inferior performance of NNR-GRRLF compared with GM-GRRLF can be caused by 1) the regularization parameter setting needs further refining; 2) part of the covariate signal may have been misinterpreted as the latent signal.

In Table 1 we present the computational cost for the above methods. We notice that despite NNR having a much more elegant formulation, it is computationally much more costly than the other alternatives (it takes roughly six CPU hours while all others take less than 1.5 seconds). This is because there is no direct correspondence between the rank and nuclear norm, thus one has to traverse the parameter space to identify the optimal parameter setting, via the costly GCV procedure. Smarter parameter space traversing strategies may significantly cut the cost, but it still takes tens of seconds to compute the generalization error for a fixed parameter pair — still more expensive than other methods[2]. The redundant parameterization of NNR-GRRLF also drags its efficiency and makes it less scalable than GM-GRRLF. We note that there are a few nuclear norm regularization optimization algorithms that are more efficient compared with the Jaggi-Hazan algorithm (Avron et al., 2012; Zhang et al., 2012b; Mishra et al., 2013; Chen et al., 2013; Hsieh and Olsen, 2014); however, these algorithms are mostly specific to certain problems and thus can not be easily extended to solve GRRLF. We therefore leave the topic of more efficient NNR-GRRLF optimization for future research, and we present some discussions on a few possible directions

[2]The computation time is also very much dependent on the stopping criteria, and therefore some compromises in the solution accuracy can also reduce the cost.

Table 1: Computation time for different methods

Method	LSR	RRR	GM-GRRLF	NNR-GRRLF
Time	< 0.01 s	< 0.01 s	1.20 s	2.09×10^4 s

in Appendix D. In the following experiments, we will exclude NNR-GRRLF due to its excessive computational burden.

To better see how this robustness can improve the estimate and in turn boost sensitivity, we varied the intensity of the covariate response and the latent response. For the covariate response experiment, we benchmarked the performance under the null, low SNR and high SNR cases, where B is scaled by 0, 0.1 and 1 respectively while fixing other settings. For the latent response experiment, we similarly test the none, weak and strong latent influence via scaling V by 0, 0.5 and 2, which accounts for $0\%, 16.7\%$ and 73.3% of the total variance respectively. All experiments were repeated for $m = 500$ times to ensure stability and for the sensitivity study we ran $m_{\text{perm}} = 100$ permutations to empirically estimate p-values, further details are provided in the Appendix. The results are summarized in Figure 8. Figure 8(a,c,d) gives the p-value distributions from the sensitivity experiment. The distribution of p-values from all three methods fall within expected region for the null case in Figure 8(a), confirming the validity of the permutation procedure. Figure 8(c-d) show that GRRLF significantly improves the sensitivity over RRR and LSR. Figure 8(b) provides a box plot of the squared difference between the estimated response and expected response on a log scale for different latent response intensities. In all cases GRRLF gives the best estimate followed by RRR. It is interesting to observe that while RRR gives a better parameter estimate compared with LSR, the latter appears to be more sensitive in our experiments.

3.2 Real-world data

736 Caucasian subjects with both genetic and *tensor based morphometry* (TBM) data from ADNI1 (`http://adni.loni.usc.edu`) are used in the current analysis. Similar to previous investigations (Stein et al., 2010a; Hibar et al., 2011; Ge et al., 2012), only age and gender are included

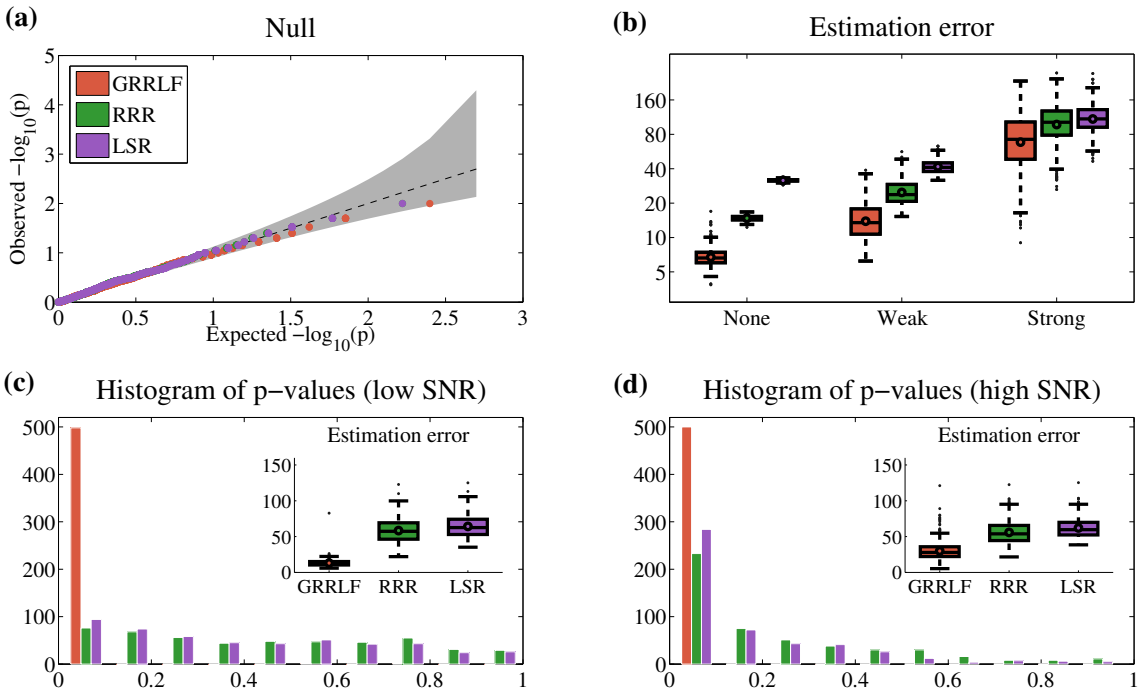


Figure 8: (a) $\log_{10}$ P-P plot of p-values under *null* model, shaded region corresponds to the 95% confidence interval under *null*. (b) Box plot of estimation error with different latent intensity. (c) Histogram of p-values and box plot of estimation error for low SNR case.(d) Histogram of p-values and box plot of estimation error for high SNR case. GRRLF demonstrates improved sensitivity and reduced estimation error compared to its commonly used alternatives under various experimental setups.

as covariates. We use LS-PCA to estimate the dimensionality of the latent space and then alternate between least square and PCA to decompose the image Y into the covariate component C, the latent component L and the residual component R, *i.e.* $Y = C + L + R$. We call $J = R + L$ the joint component. We have chosen the LS-PCA implementation to demonstrate because this is the simplest form of GRRLF, computationally efficient and there is no parameter to be tuned, which makes it more likely to be used in practice compared with other more sophisticated implementations. Then we apply the LSKM to estimate the gene-wise genetic effect on J, L and R respectively for each voxel. A total of $26,664$ genes and $29,479$ voxels enter the study. We thresholded the significance image with threshold $p < 10^{-3}$ and use the largest cluster size (in RESEL units) as the test statistic. All p-values, including those of the voxel-level LSKM test score and the largest cluster size statistics, were determined via nonparametric permutations. As a *post hoc* validation step, we searched the *Genevisible* database (Nebion, 2014; Hruz et al., 2008) for the top genes identified in each category to examine whether they are highly expressed in neuron-related tissues (HENT) [3]. Consult Appendix F for more details on the study sample, data preprocessing and statistical analyses. The latent factor identification results are visualized in Figure 9 and the GWAS results are tabulated in Table 2.

Figure 9(a) indicates that the first three eigen-components are the dominant parts of J, and thus we identify them as the latent components, i.e. $t = 3$. Figure 9(b) gives the spatial maps of the decomposed latent components, and interestingly they seem to respectively correspond to white matter, ventricles and gray matter. For the GWAS analysis, smaller p-values are obtained for the top hits in factorized analyses. While no gene from the above three analyses survived stringent Bonferroni correction, three of the genes, all from the factorized GWAS analyses, survived the FDR significance level $q = 0.2$ suggested by Efron (2010). More than half of the top entries identified in the factorized analyses have been reported to be relevant in neuronal researches, indicating that the results from the factorized analyses are biologically relevant.

The top hit in Table 2 is *CACNA1C* (overlapping with *DCP1B*), an L-type voltage gated calcium

[3] Neuron-related tissues are defined as neuronal cells or brain tissues. YES: neuron-related tissues among the top 5 out of 381 tissue types in terms of expression level, NO: otherwise, N/A: information not available for the gene.

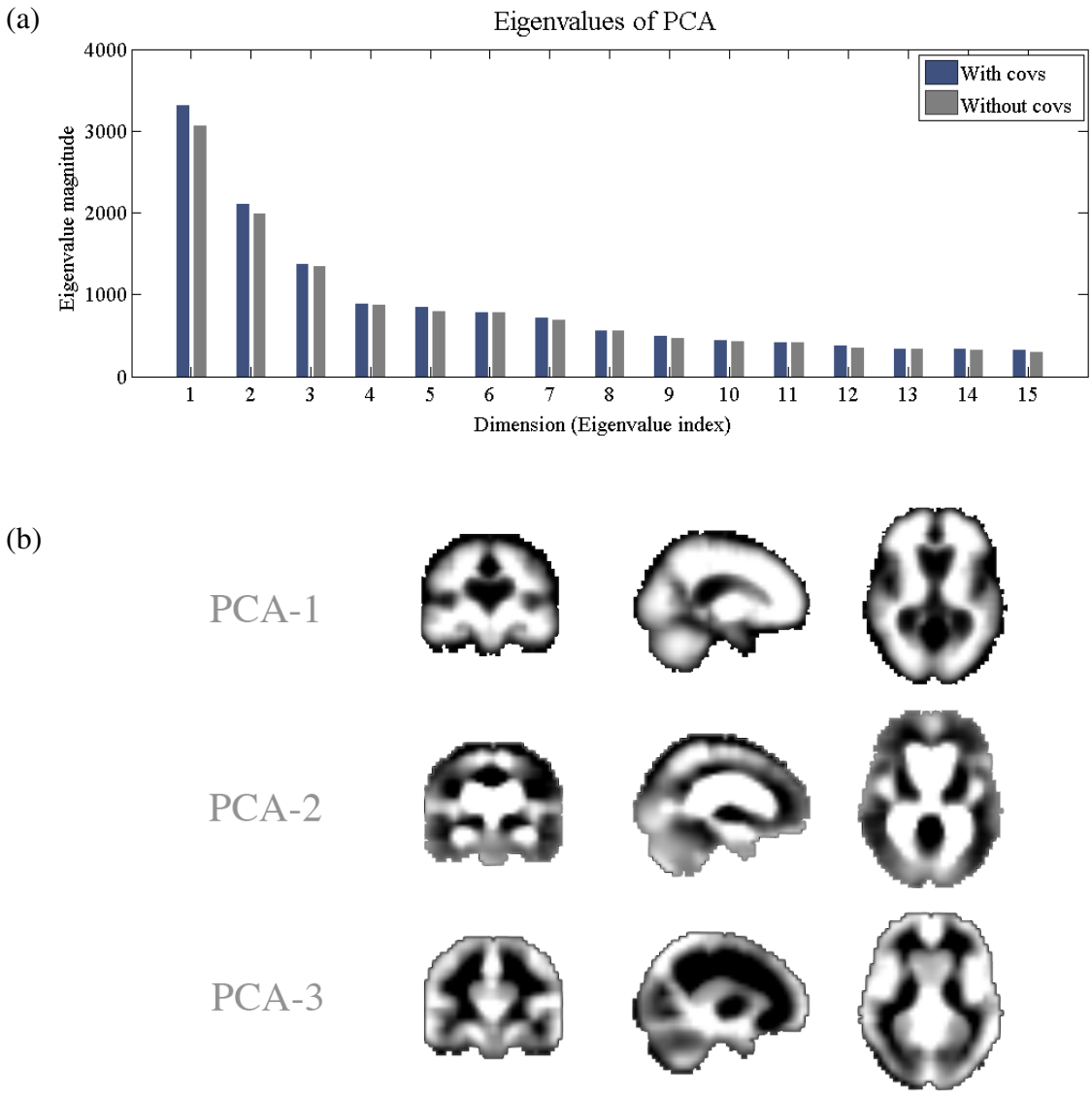


Figure 9: (a) Eigenvalues from PCA with or without the covariates. (b) Spatial maps of the first three latent factors. First three eigencomponents encodes significantly more variances compares with other eigencomponents thus being identified as the latent components.

Table 2: GWAS results

Joint component J								
Chr.	Gene Name	SNPs	Cluster Size (in RESEL)	p_{uc}	q_{fdr}	Nearby Genes	HENT	Related Functions
1	PGM1	16	1.89	7.69E-05	7.66E-01		NO	AD
11	CCDC34	8	1.72	1.20E-04	7.66E-01	BDNF	NO	psychiatric disorder risk factor*
14	CTD-2555O16.1	3	1.59	1.56E-04	7.66E-01		N/A	
14	TEX21P	5	1.51	2.16E-04	7.66E-01		N/A	
6	EFHC1	16	1.5	2.21E-04	7.66E-01		NO	neuroblasts migration, epilepsy
12	RP11-421F16.3	2	1.42	3.45E-04	7.66E-01	TM7SF3	NO	
2	CHRNA1	3	1.42	3.49E-04	7.66E-01		NO	autism
3	MARK2P14	2	1.42	3.51E-04	7.66E-01		NO	
16	RP11-488I20.8	1	1.37	4.31E-04	7.66E-01	LINC01566	YES	
15	ISLR	1	1.36	4.31E-04	7.66E-01		NO	
Latent component L								
Chr.	Gene Name	SNPs	Cluster Size (in RESEL)	p_{uc}	q_{fdr}	Nearby Genes	HENT	Related Functions
12	**GRIN2B**	**179**	**3.99**	**7.50E-06**	**2.00E-01**		**YES**	**learning, memory, AD, etc.**
7	AC074389.7	1	3.09	2.63E-05	3.50E-01	ELFN1	NO	seizure, ADHD
8	HPYR1	1	2.26	4.50E-05	4.00E-01		YES	
7	ELFN1	6	1.62	6.75E-05	4.40E-01		NO	seizure, ADHD
12	RSRC2	4	1.38	8.25E-05	4.40E-01		NO	
2	CHRNA1	3	1.06	1.14E-04	4.93E-01		NO	autism
17	TAC4	2	0.85	1.29E-04	4.93E-01		NO	
11	RP11-872D17.8	9	0.4	2.63E-04	7.81E-01	PRG2/3,SLC43A3	NO	
1	EMC1	8	0.35	4.63E-04	7.81E-01		YES	
17	AP2B1	27	0.35	4.63E-04	7.81E-01		NO	schizophrenia
Residual component R								
Chr.	Gene Name	SNPs	Cluster Size (in RESEL)	p_{uc}	q_{fdr}	Nearby Genes	HENT	Related Functions
12	**RP5-1096D14.6**	**2**	**2.51**	**9.38E-06**	**1.25E-01**	**CACNA1C**	**YES**	**psychiatric disease risk factor***
12	**DCP1B**	**12**	**2.48**	**9.38E-06**	**1.25E-01**	**CACNA1C**	**NO**	**psychiatric disease risk factor***
15	PYGO1	10	1.79	8.06E-05	5.36E-01		NO	wnt signaling pathway, AD
11	DOC2GP	1	1.72	1.05E-04	5.36E-01		N/A	
6	GLYATL3	7	1.71	1.05E-04	5.36E-01		N/A	
2	MREG	36	1.63	1.41E-04	5.36E-01		NO	
4	ZGRF1	9	1.62	1.41E-04	5.36E-01		NO	
12	FAM216A	2	1.58	1.73E-04	5.75E-01		YES	neurodegenerative disease
10	CEP164P1	16	1.52	2.44E-04	7.22E-01		N/A	
10	RP11-285G1.15	22	1.47	3.13E-04	8.28E-01	RSU1P2	YES	substance addiction†

GWAS results, showing the distinct findings between joint, latent and residual component. "Nearby Genes", those genes that lie in close vicinity (within a few hundred KB) of the primary gene that has showed significant association, in some cases the genes are co-located so the nearby genes can also be regarded as the primary gene; "HENT", the primary gene (or the nearby gene if such information is not available for the primary gene) is highly expressed in neuron-related tissues (see main text for detailed definition); *, the function is related to the nearby gene(s); †, the function is related to the functioning gene. We have highlighted genes that are statistical significant after multiple comparison and underlined genes of particular interest.

channel subunit gene well known for its psychiatric disease susceptibility (PGC et al., 2013). The significance map between *CACNA1C* and the TBM map is overlaid on the population template in Figure 10, and it can be seen that the voxels susceptible to this influence are clustered within the orbitofrontal cortex, and overlapping *gyrus rectus* and *olfactory* regions, which include the caudal orbitofrontal cortex Brodmann area 13 (Öngür et al., 2003). We further conducted SNP-wise association for all the imputed SNPs within 500 KB of *CACNA1C*'s coding region. Only SNPs that have a minor allele frequency over 0.1 are included. The result is presented in Figure 11. The peak association is achieved at SNP rs2470446 (maf=0.47), which is imputed. For the genotyped SNPs, rs2240610 (maf=0.49) yields the largest association. No association is observed between rs2240610 and the Alzheimer or dementia diagnostic state of the subjects (all $p > 0.05$). In the following we use *DCP1B* as a surrogate for *CACNA1C* as the majority of *CACNA1C* SNPs lie outside the genetic hot spot. We extract the first eigen-component of the largest voxel cluster associated with *CACNA1C* and plot them against the genotype of SNP rs2240610. Subjects with genotype 'AA' have significantly different responses compared with the other two genotypes (t-test, $p = 8.55 \times 10^{-10}$), which have similar responses compared with each other (t-test, $p = 0.50$). This result suggests that the recessive model is appropriate for the genetic effect. A similar distribution is observed for the mean response of the cluster.

4 Discussion

In this paper, we propose a general framework of reduced rank latent factor regression for neuroimaging applications. In summary, we (1) reduce the variance of the covariate effect estimate by simultaneously (a) projecting the predictors onto a lower dimensional effective subspace and (b) conditioning on the latent components that are dynamically estimated; (2) we use additional constraints such as smoothness of the response field to regularize the solution; (3) we recast the problem into a sequence of block-manifold optimization problems and effectively solve them by *Riemannian manifold optimization*; (4) we present an alternative nuclear norm regularization based

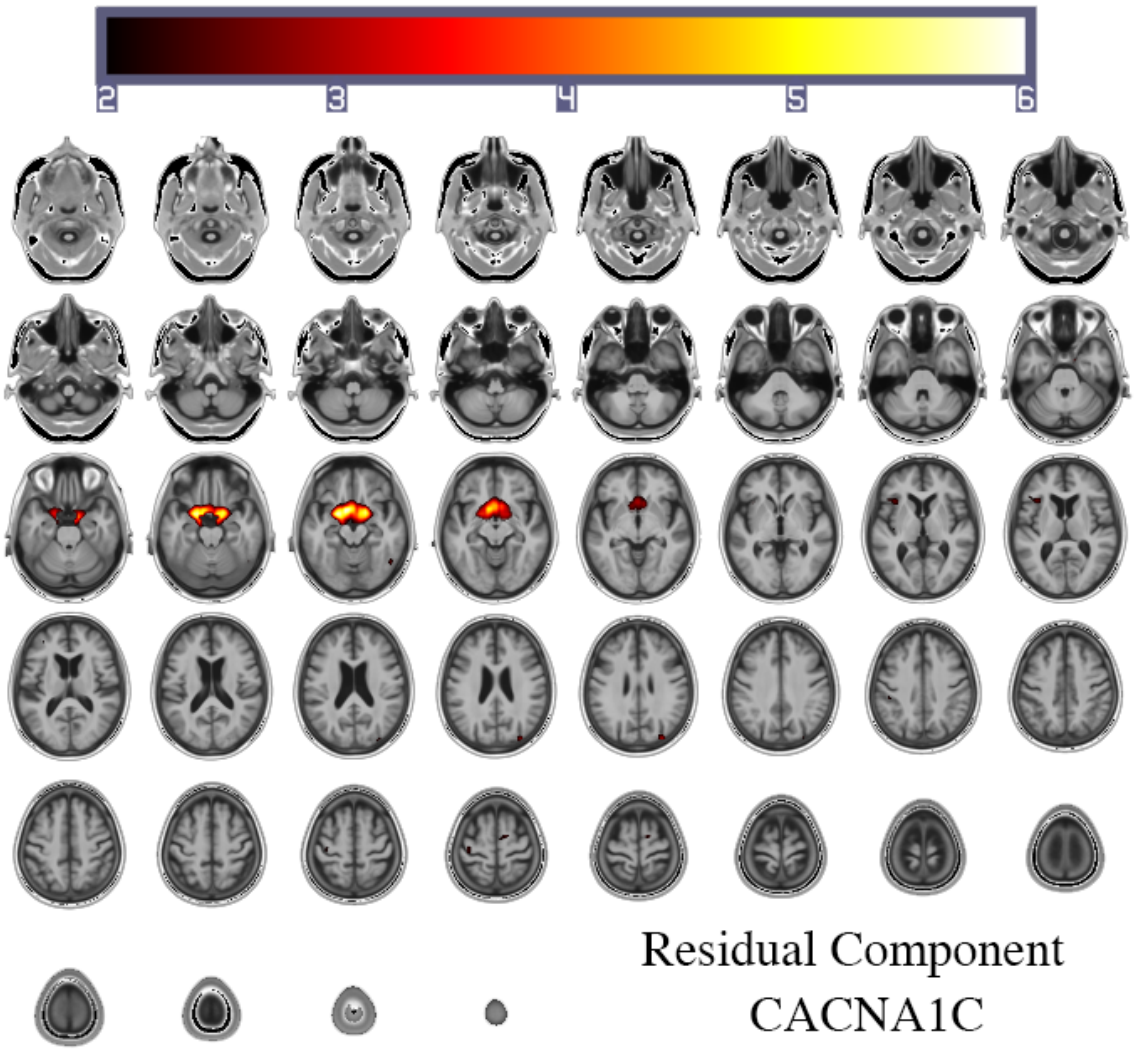


Figure 10: Significance map of *CACNA1C*, color coded with $-log_{10}(p)$. Voxels susceptible to the genetic influence from *CACNA1C* are clustered within the orbitofrontal cortex

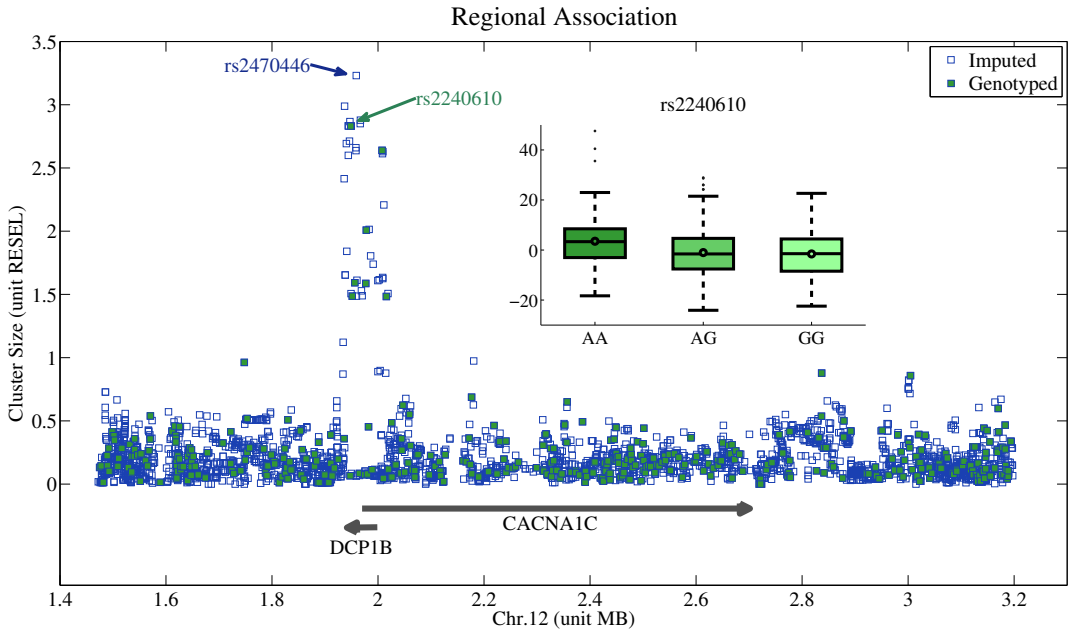


Figure 11: Regional SNP-wise cluster size analyses for *CACNA1C*. Inset: distribution of first principal coordinate of the voxels within the largest cluster according to the genotype of SNP rs2240610.

formulation of GRRLF with which the global optimum can be achieved; (5) we present a least squares kernel machines based procedure for brain-wide GWAS conditioning on the latent factors.

Our method exploits the structured nature of the imaging data to better factorize the signal observed. The application of our method to a real-world dataset suggests that this factorization improves upon the sensitivity over existing brain-wide GWAS methods and gives biologically plausible results. The most significant gene identified, *CACNA1C*, is a widely recognized multi-spectrum psychiatric risk factor and has been intensively studied. Our result lends further evidence for the pleiotropic role it plays. Most of the top genes that we identified are found to be either relevant to psychiatric diseases or highly expressed in neuronal tissues, lending plausibility to our framework.

4.1 Methodology assessment

Our method reports two genes surviving the FDR threshold at $q = 0.2$ while previous work has not fund any (Hibar et al., 2011). We note our imaging-genetic solution is closely related to Ge

et al. (2012) where they use analytical approximation of LSKM statistics, *extreme value theory* (EVT) and *random field theory* (RFT) to make inferences. Different from Ge et al. (2012), our null simulations fail to support the use of these analytical approximations, so only permutation-based results are reported in the current study. None of the top genes identified from the residual component study have been reported in Ge et al. (2012) and Hibar et al. (2011) while there are a few overlaps for the genes from the joint and latent component study. This suggests that conditioning for the hidden variables might be important to reveal certain otherwise buried signals.

While GRRLF can be implemented in various forms, the key idea underlying our framework is three-fold: 1) using the structure of brain imaging to estimate the latent components; 2) conditioning on the latent component to reduce the variance of covariate effect of interest; 3) estimating the effective dimensionality of the covariate further reduces variance. An interesting comparison can be made with the *linear mixed model* (LMM) which has recently gained popularity in GWAS studies (Eu-ahsunthornwattana et al., 2014), where a kinship matrix, estimated from either pedigree or genome sequences, is used to structure the covariance matrix for genetic random effects. LMM deals with univariate response so it can only look into the kinship matrix for structured unexplained variance, while for neuroimaging data the richness of the structural information allows further decomposition of the observed signals. Current large scale multi-center neuroimaging collaborations often use comprehensive survey to capture as much population variance as possible and researchers are compelled to include more predictors in their model to factor out the variances in the data. However, the price paid is the *degrees of freedom* (DOF) and therefore more uncertainty in estimating the effect of interest. Enforcing proper regularizations, in our case constraining the effective dimensions of the predictors, serves to balance the trade-off between the explained variance and DOF.

The three sets of results just presented show that each each decomposition scheme has its advantages and that they are complementary to each other. The residual component approach is more sensitive to weak signals that would otherwise be dominated by a large latent component effect. The latent component approach has the advantage that it acts to reduce noise, but may

not detect local effects. The joint component approach is useful if there are contributions of both global and local effects. We therefore suggest that the results with all three approaches should be compared with each dataset analyzed.

4.2 Biological significance

CACNA1C is known as one of the risk genes for a wide spectrum of psychiatric disorders including bipolar disease, schizophrenia, and major depression and autism. Its association with susceptibility to psychiatric disorders has been consistently confirmed by several large-scale genome-wide association studies (PGC et al., 2013) thus making it one of the most replicable results in psychiatric genetics. A series of human brain imaging and behavioral studies have shown morphological and functional alterations in individuals carrying the *CACNA1C* risk allele on a macroscopic level (Bigos et al., 2010; Franke et al., 2010; Zhang et al., 2012a; Tesli et al., 2013; Erk et al., 2014), and it has been experimentally confirmed that the risk variant will also affect cellular level electrophysiology using induced human neuron cell lines (Yoshimizu et al., 2014). An Australian twin study has previously reported *CACNA1C* to be significantly associated with white matter integrity and function as a hub in the expression network belonging to the enriched gene ontology category "synapse" (Chiang et al., 2012). Previous studies on the ADNI dataset have also reported significant genetic interactions for *CACNA1C* using Positron Emission Tomography (PET) imaging (Koran et al., 2014) and LASSO screening with candidate phenotypes (Yang et al., 2015), which all involve certain *'conditioning'* for the contribution from *CACNA1C* to be detected. Animal AD models have confirmed several results from human studies (Hopp et al., 2014) and related pathways have been identified as a therapeutic target (Liang and Wei, 2015) for AD. Interestingly, a recent multi-site large-scale voxel level functional connectivity study, which included 939 subjects, has revealed that functional connectivity patterns in the orbitofrontal cortex region are significantly altered in depression patients (Cheng et al., 2015). Also gray matter volume reductions are reported in the same area in depression patients (Ballmaier et al., 2014). The results from these studies are consistent with the assumption that *CACNA1C* affects depression susceptibility through the

orbitofrontal region, a hypothesis to be tested in future studies.

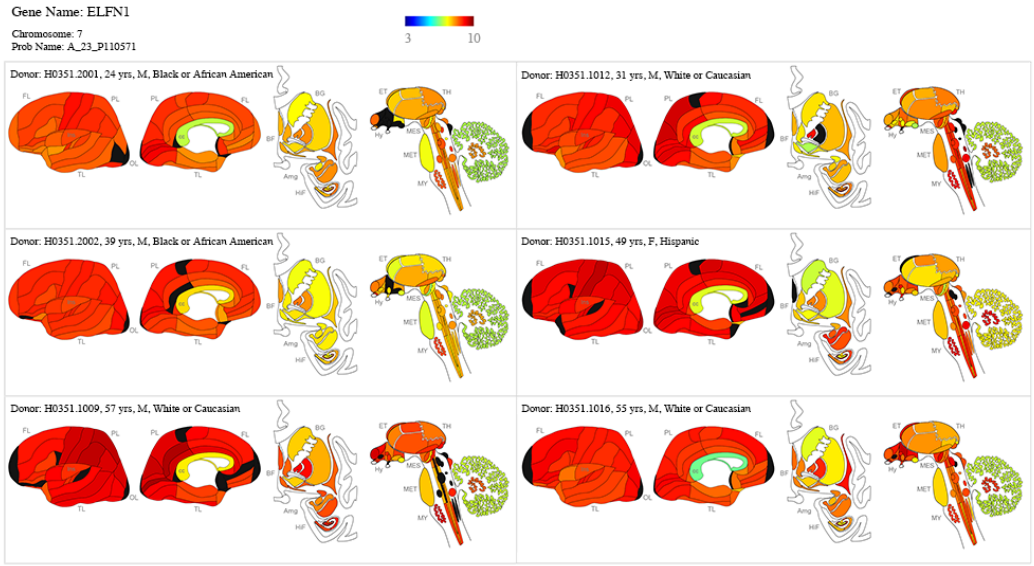
ELFN1 has been implicated to be associated with seizures and ADHD in both human clinical samples and animal models (Tomioka et al., 2014; Dolan and Mitchell, 2013). The expression of *ELFN1* localizes mostly to excitatory postsynaptic sites (Sylwestrak and Ghosh, 2012) and recent studies show that the *ELFN1* gene specifically controls short-term plasticity, which denotes changes in synaptic strength that last up to tens of seconds, at some synapse types (Blackman et al., 2013). Data from the *Allen's Brain Atlas* (Hawrylycz et al. (2012); http://human.brain-map.org) also show that ELFN1 is highly expressed in the cortical regions (Figure 12(a)), consistent with the *ELFN1* significance map we obtained from the ADNI dataset (Figure 12(b)).

GRIN2B encodes the N-methyl-D-aspartate (NMDA) glutamate receptor *NR2B* subunit and is well known to be involved in learning and memory (Tang et al., 1999), structural plasticity of the brain (Lamprecht and LeDoux, 2004) and excitotoxic cell death (Parsons et al., 2007), and has age-dependent prevalence in the synapse (Yashiro and Philpot, 2008). Therefore, the relationship between *NR2B* subunit gene *GRIN2B* variants and AD has attracted a large amount of attention and interest. Many studies have confirmed that the *NR2B* subunit is down-regulated significantly in susceptible regions of AD brains (Bi and Sze, 2002; Farber et al., 1997; Hynd et al., 2004). Actually *GRIN2B* is already a therapeutic target in Alzheimer's disease and has also been indicated by several studies in the literature (Jiang and Jia, 2009; Stein et al., 2010b).

The gene *PGM1* encodes the protein Phosphoglucomutase-1. The level of this enzyme was found to be significantly altered in the hippocampus of patients who suffer from AD compared with control hippocampus using two-dimensional gel electrophoresis and mass spectrometry techniques (Sultana et al., 2007). Down-regulation of this gene might have an effect on memory and cognitive functions in human brains.

The pygopus gene of Drosophila encodes an essential component of the Armadillo (β-catenin) transcription factor complex of canonical *wnt* signaling (Schwab et al., 2007). The *wnt* signaling pathway has been implicated in a wide spectrum of physiological processes during the development

(a)



(b)

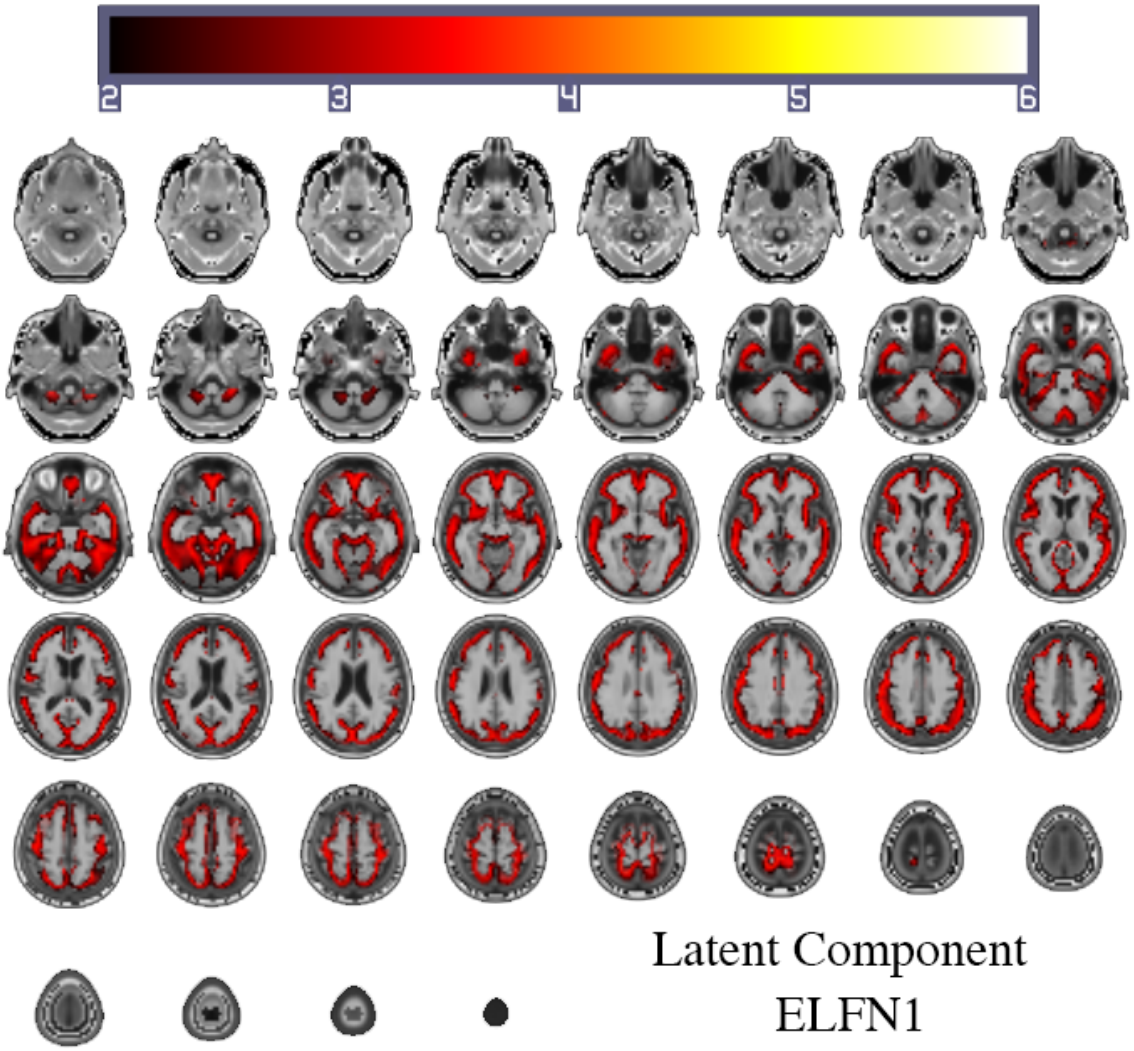


Figure 12: (a) *ELFN1* expression profile from Allen Brain Atlas. (b) Significance map of *ELFN1*, color coded with $-log_{10}(p)$. Cortical regions show elevated *ELFN1* expression and they are also under the genetic influence of the same gene.

of the central nervous system and assumes some roles in mature synapses that could cause cognitive deficiencies (Oliva et al., 2013). A recent study has pointed out that aberrant *wnt* signaling pathway function is associated with medial temporal lobe structures of Alzheimer's disease and *PYGO1* is differentially expressed in an AD population using post-mortem brain samples (Riise et al., 2015).

FAM216A has been reported to be a risk gene for neurodegenerative diseases from an integrated multi-cohort transcriptional meta-analysis study using $1,270$ post-mortem central nervous system tissue samples (Li et al., 2014). *AP2B1* is reported to be differentially expressed in a rat model for schizophrenia (Zhou et al., 2010). *CHRNA* family genes have been implicated as susceptible targets in autism spectrum disorders (Lee et al., 2012). *EFHC1* mutations are known to cause juvenile myoclonic epilepsy (Suzuki et al., 2004; Stogmann et al., 2006; Noebels et al., 2012). *CCDC34* has been previously reported to be associated with ADHD and autism (Shinawi et al., 2011) and it locates next to the gene *BDNF* which is known to be a risk factor for psychiatric disorders (Petryshen et al., 2010). *RSU1P2* is the pseudogene of *RSU1*, *i.e.* a DNA sequence that is similar to the functioning gene *RSU1* but nonetheless unable to produce functional protein products and only assuming regulatory roles. It is reported that RSU1 has a conserved role regulating reward related phenotypes such as ethanol consumption, ranging from *Drosophila* to humans (Ojelade et al., 2015).

4.3 Future directions

The current study provides for advances in a number of directions. On the biological side, a few interesting assumptions have been made combining the results from the ADNI dataset and existing studies. These assumptions can be checked on the data from other phases of the ADNI project, for example ADNI GO and ADNI2, or other population samples. The proposed method can also be applied to longitudinal recordings from the ADNI dataset, where brain-wide genome-wide imaging-genetic investigations are rare due to the fact that the observed phenotype is a function of time, *i.e.* a one dimensional tensor, thus unsuited for most neuroimaging-genetic solutions.

On the methodological side, many aspects can be further improved in the future. For example,

we do not deal with the identifiability issue of the model to maximize the generality of the formulation. More stringent constraints are expected to theoretically ensure the identifiability under certain assumptions, which is left for future investigations. Pragmatically it will be interesting to compare the empirical performance of GRRLF with latent factors estimated from different models, say ICA. More computationally efficient estimation procedures, and more sensitive yet less expensive statistical tests are also important topics for future exploration.

Acknowledgements

The authors report no conflict of interest. The authors would like to thank the two reviewers and the editor for their insightful comments, especially for bringing the NNR to our attention. CY Tao is supported by the China Scholarship Council (CSC) and National Natural Science Foundation of China (No. 11101429 and No. 11471081). TE Nichols is supported by the Wellcome Trust (100309/Z/12/Z) and NIH R01 EB015611-01. JF Feng is a Royal Society Wolfson Research Merit Award holder and he is also partially supported by the National High Technology Research and Development Program of China (No. 2015AA020507) and the Key Project of Shanghai Science & Technology Innovation Plan (No. 15JC1400101). The research is partially supported by the National Centre for Mathematics and Interdisciplinary Sciences (NCMIS) of the Chinese Academy of Sciences and Key Program of the National Natural Science Foundation of China (No. 91230201), and the Shanghai Soft Science Research Program (No. 15692106604). Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimers Association; Alzheimers Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche

Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. The support and resources from the Apocrita HPC system at Queen Mary University of London are gratefully acknowledged. The authors would also like to thank Prof. CL Leng, Prof. G Schumann, Dr. T Ge, Dr. S Desrivières, Dr. TY Jia, Dr. L Zhao, Dr. W Cheng and Dr. B Xu for fruitful discussions.

Appendix A Connection to other models

Different choices of loss function ℓ, covariate dimension d and latent dimension t of (5) give different commonly used statistical models. In Table S1 $\|\cdot\|_{\text{Fro}}$ denotes the Frobenius norm, $\eta(\cdot), \tau(\cdot)$ and $\alpha(\cdot)$ are the functions charactering the exponential family distributions (McCullagh and Nelder, 1989), $\Upsilon(\cdot)$ certain dependency measure (Suzuki and Sugiyama, 2013; Gretton et al., 2005), $\vartheta(\cdot)$ certain independence measure (Bach and Jordan, 2003; Hyvärinen et al., 2004; Smith et al., 2012), $\Psi(\cdot)$ the basis functions of some functional space (Ramsay and Silverman, 2005) and $\{\lambda_i\}$ the regularization parameters (Zou and Hastie, 2005).

Table S1: Connection to other models

ℓ	d	t	Statistical model
$\|Y - XB\Phi\|_{\text{Fro}}^2$	$< p$	0	Reduced rank regression
$-\Upsilon(Y, XB\Phi)$	$< p$	0	Supervised dimension reduction
$-\eta(\Phi) \cdot \tau(Y, X) + \alpha(\Phi)$	p	0	GLM (exponential family)
$\|Y - X\Phi\|_{\text{Fro}}^2 + \lambda_1\|\Phi\|_1 + \lambda_2\|\Phi\|_2^2$	p	0	Lasso / Elastic net
$\|Y - \Psi(X)B\Phi\|_{\text{Fro}}^2$	$< \lvert\Psi\rvert$	0	Functional PCA
$\|Y - L\Gamma\|_{\text{Fro}}^2$	0	any	Generalized PCA
$-\vartheta(Y\Gamma^\top)$	0	any	ICA

Appendix B Reduced rank regression

Here we detail the implementation of reduced rank regression (Izenman, 1975). Assume that X and Y have been demeaned. For $Y \in \mathbb{R}^q$, $X \in \mathbb{R}^p$, rank-d reduced rank regression intends to find the $A \in \mathbb{R}^{q\times d}, B \in \mathbb{R}^{d\times p}$ that minimizes $\mathbb{E}[\|Y - ABX\|_2^2]$. Denote Σ_{XX} the covariance matrix of X and Σ_{YX} the cross-covariance matrix between Y and X. Denote $\Sigma = \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{YX}^\top$ and $\Sigma = V\Lambda V^\top$ its eigen-decomposition, where V is a unitary matrix and Λ a diagonal matrix with non-negative entries in descending order. Denote $V_d \in \mathbb{R}^{q\times d}$ the first d columns of V, then the solution of reduced-rank regression with rank d is

$$A^* = V_d, B^* = V_d^\top \Sigma_{YX}\Sigma_{XX}^{-1}.$$

Appendix C Nuclear norm regularized GRRLF

First we prove solving (8) is a convex optimization problem.

Lemma 1 (Jaggi et al. (2010), Lemma 1)**.** *For any non-zero matrix $\boldsymbol{X} \in \mathbb{R}^{n\times m}$ and $t \in \mathbb{R}$, $\|\boldsymbol{X}\|_* \le \frac{t}{2}$ iff there exists $\boldsymbol{A} \in \mathbb{S}_{PSD}^n$ and $\boldsymbol{B} \in \mathbb{S}_{PSD}^m$, such that $\begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{B} \end{pmatrix} \ge 0$ and $tr(\boldsymbol{A}) + tr(\boldsymbol{B}) = t$.*

Lemma 2 (Laurent and Vallentin (2012))**.** *$\mathbb{S}_{PSD}^n$ is a cone.*

Lemma 3. *Define $\mathbb{S}_{PSD}^n(t) := \{\boldsymbol{A} \in \mathbb{S}_{PSD}^n | tr(\boldsymbol{A}) = t\}$, then $\mathbb{S}_{PSD}^n(t)$ is a convex set.*

Proof. For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}^n_{\text{PSD}}(t)$ and $\omega \in [0,1]$, for $\boldsymbol{C} = \omega\boldsymbol{A} + (1-\omega)\boldsymbol{B}$, we have

$$\text{tr}(\boldsymbol{C}) = \text{tr}(\omega\boldsymbol{A} + (1-\omega)\boldsymbol{B}) = \omega\text{tr}(\boldsymbol{A}) + (1-\omega)\text{tr}(\boldsymbol{B}) = t,$$

therefore $\boldsymbol{C} \in \mathbb{S}^n_{\text{PSD}}(t)$, which concludes the proof. □

Theorem 1. *$f(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K)$ is a convex function, where $\boldsymbol{X}_k \in \mathbb{R}^{n_k \times m_k}, k \in [K]$. let $\{t_k | k = [K]\}$ be a set of positive real numbers, then the nuclear norm regularized problem*

$$(\boldsymbol{X}_1^*, \cdots, \boldsymbol{X}_K^*) = \underset{\|\boldsymbol{X}_k\|_* \le t_k/2}{\arg\min}\ f(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K) \tag{10}$$

is equivalent to the convex problem

$$(\boldsymbol{Z}_1^*, \cdots, \boldsymbol{Z}_K^*) = \underset{\boldsymbol{Z}_k \in \mathbb{S}^{n_k+m_k}_{PSD}(t_k)}{\arg\min}\ \widetilde{f}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_K), \tag{11}$$

where $\boldsymbol{Z}_k = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{X} \\ \boldsymbol{X}^\top & \boldsymbol{B} \end{pmatrix}$ *and* $\widetilde{f}(\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_K) := f(\boldsymbol{X}_1, \cdots, \boldsymbol{X}_K)$.

Proof. Since the Cartesian product of convex sets is also a convex set, so by Lemma 3 we know $\prod_{\otimes} \mathbb{S}^{n_k \times m_k}_{\text{PSD}}(t_k)$ is a convex set. And the convexity of $\widetilde{f}$ is inherited from f. The proof is completed by applying Lemma 1 to obtain the bound of $\|\boldsymbol{X}_k\|_*, k \in [K]$. □

Setting $K = 2$ in Theorem 1 proves the convexity of (8).

Now we elaborate how to efficiently compute the solutions given the NN constraint t_1 and t_2. First we extend $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{L}}$ to $\boldsymbol{Z}_B = \begin{pmatrix} \boldsymbol{A} & \widetilde{\boldsymbol{B}} \\ \widetilde{\boldsymbol{B}}^\top & \boldsymbol{C} \end{pmatrix}$ and $\boldsymbol{Z}_L = \begin{pmatrix} \boldsymbol{C} & \widetilde{\boldsymbol{L}} \\ \widetilde{\boldsymbol{L}}^\top & \boldsymbol{D} \end{pmatrix}$, where $\boldsymbol{Z}_B \in \mathbb{S}^{(p+m)}_{\text{PSD}}(t_1), \boldsymbol{Z}_L \in \mathbb{S}^{n+m}_{\text{PSD}}(t_2)$. Let $\widetilde{\boldsymbol{Z}}_B = t_1^{-1}\boldsymbol{Z}_B, \widetilde{\boldsymbol{Z}}_L = t_2^{-1}\boldsymbol{Z}_L$, $\boldsymbol{X}_t = t_1\boldsymbol{X}$, $\boldsymbol{B}_t = t_1^{-1}\widetilde{\boldsymbol{B}}$, $\boldsymbol{L}_t = t_2^{-1}\widetilde{\boldsymbol{L}}$, define

$$\widetilde{f}(\widetilde{\boldsymbol{Z}}_B, \widetilde{\boldsymbol{Z}}_L) = f_t(\boldsymbol{B}_t, \boldsymbol{L}_t, t_2) = \|\boldsymbol{Y} - \boldsymbol{X}_t\boldsymbol{B}_t\boldsymbol{H} - t_2\boldsymbol{L}_t\boldsymbol{H}\|_F^2 = f(\boldsymbol{B}, \boldsymbol{L}).$$

Notice $\text{tr}\left(\widetilde{\boldsymbol{Z}}_B\right) = \text{tr}\left(\widetilde{\boldsymbol{Z}}_L\right) = 1$, so we can use the Hanzan algorithm (Hazan, 2008) to opti-

mize $\widetilde{f}(\widetilde{\boldsymbol{Z}}_B, \widetilde{\boldsymbol{Z}}_L)$ over $\mathbb{S}_{BL} \coloneqq \mathbb{S}^{p+m}_{\text{PSD}}(1) \times \mathbb{S}^{n+m}_{\text{PSD}}(1)$. Let $\boldsymbol{R}_t = \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}_t\boldsymbol{H} - t_2\boldsymbol{L}_t\boldsymbol{H}$, we have $\nabla_{\boldsymbol{B}_t} f_t = 2\boldsymbol{X}^\top\boldsymbol{R}_t\boldsymbol{H}^\top, \nabla_{\boldsymbol{L}_t} f_t = 2t_2\boldsymbol{R}_t\boldsymbol{H}^\top$. Let $\boldsymbol{v}_B = \text{MaxEV}(-\nabla_{\boldsymbol{B}_t} f_t)$ and $\boldsymbol{v}_L = \text{MaxEV}(\nabla_{\boldsymbol{L}_t})$, where $\text{MaxEV}(\boldsymbol{A})$ computes the eigenvector corresponds to the maximal eigenvalue of $\boldsymbol{A} \in \mathbb{S}^n_{\text{PSD}}$; by Hazan algorithm $\Delta\boldsymbol{B}_t \coloneqq \boldsymbol{v}_B\boldsymbol{v}_B^\top - \boldsymbol{B}_t$ and $\Delta\boldsymbol{L}_t \coloneqq \boldsymbol{v}_L\boldsymbol{v}_L^\top - \boldsymbol{L}_t$ are the search directions for $\widetilde{\boldsymbol{Z}}_B$ and $\widetilde{\boldsymbol{Z}}_L$ respectively. By minimizing $\widetilde{f}_t(\boldsymbol{B}_t, \boldsymbol{L}_t, \alpha) \coloneqq f(\boldsymbol{B}_t + \alpha\Delta\boldsymbol{B}_t, \boldsymbol{L}_t + \alpha\Delta\boldsymbol{L}_t)$ with respect to learning rate $alpha$, we have the optimal learning rate given by

$$\alpha^* = \frac{\langle \boldsymbol{M}_B, \Delta\boldsymbol{B}_t \rangle_F + \langle \boldsymbol{M}_L, \Delta\boldsymbol{L}_t \rangle_F}{\langle \boldsymbol{N}_B, \Delta\boldsymbol{B}_t \rangle_F + \langle \boldsymbol{N}_L, \Delta\boldsymbol{L}_t \rangle_F}, \tag{12}$$

where

$$\begin{aligned}
\boldsymbol{D}_{BL} &= (\boldsymbol{X}_t\Delta\boldsymbol{B}_t + t_2\Delta\boldsymbol{L}_t)\boldsymbol{\Phi}, \\
\boldsymbol{M}_B &= \boldsymbol{X}_t^\top\boldsymbol{R}_t\boldsymbol{\Phi}^\top, \boldsymbol{M}_L = t_2\boldsymbol{R}_t\boldsymbol{\Phi}^\top, \\
\boldsymbol{N}_B &= \boldsymbol{X}_t^\top\boldsymbol{D}_{BL}\boldsymbol{\Phi}^\top, \boldsymbol{N}_L = t_2\boldsymbol{D}_{BL}\boldsymbol{\Phi}^\top.
\end{aligned}$$

To further speed up the computation we adopt the "hot start" strategy, that the solution of $\left(t_1^{(1)}, t_2^{(1)}\right)$ is used to initialize the optimization for nearby parameter pair $\left(t_1^{(2)}, t_2^{(2)}\right)$.

Notice $\nabla_{\boldsymbol{Z}_B}\widetilde{f}$ and $\nabla_{\boldsymbol{Z}_L}\widetilde{f}$ are always symmetric matrices of the block form $\begin{pmatrix} \boldsymbol{0} & \boldsymbol{G} \\ \boldsymbol{G}^\top & \boldsymbol{0} \end{pmatrix}$, of which the eigenvectors are also symmetric: whenever $(\boldsymbol{v}^\top, \boldsymbol{w}^\top)^\top$ is the eigenvector for eigenvalue λ, $(\boldsymbol{v}^\top, -\boldsymbol{w}^\top)^\top$ is the eigenvector for eigenvalue $-\lambda$. Also it is easy to see $\boldsymbol{v}$ and $\boldsymbol{w}$ are the eigenvectors for $\boldsymbol{G}\boldsymbol{G}^\top$ and $\boldsymbol{G}^\top\boldsymbol{G}$ respectively, with the eigenvalue λ^2. These factor will break down the computation of $\text{MaxEV}(-\Delta_\square f_t)$ into computing the *principal eigenvectors*[4] for two lower order matrices $\boldsymbol{G}\boldsymbol{G}^\top$ and $\boldsymbol{G}^\top\boldsymbol{G}$ respectively. This can be easily achieved via Lanczos method, Ritz approximation or simply the power iteration. Further speedup can be achieved for "tall" matrices by squaring the lower order matrix product.

[4]Principal eigenvector is the eigenvector with respect to the largest eigenvalue in absolute value.

Appendix D Further discussions on NNR implementations

In this section we present some discussions on the NNR implementation to motivate more efficient algorithms. First we define the notation of SVD-thresholding operators. Consider the singular value decomposition of $\boldsymbol{Y} \in \mathbb{R}^{p\times q}$,

$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^T, \tag{13}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are respectively $p\times h$ and $q\times h$ orthonormal matrices with $h = \min(p,q)$ known as the left and right singular vectors, and diagonal matrix $\boldsymbol{D}$ consists of non-increasing non-negative diagonal elements $\boldsymbol{d}$ known as the singular values of $\boldsymbol{Y}$. For any $\lambda \geq 0$, the hard SVD-thresholding operator is defined as

$$\mathcal{H}_\lambda(\boldsymbol{Y}) = \boldsymbol{U}\mathcal{H}_\lambda(\boldsymbol{D})\boldsymbol{V}^T, \mathcal{H}_\lambda(\boldsymbol{D}) = \text{diag}(\{d_i I_{d_i>\lambda}\}), \tag{14}$$

where $I.$ is the indicator function, and the soft SVD-thresholding operator

$$\mathcal{S}_\lambda(\boldsymbol{Y}) = \boldsymbol{U}\mathcal{S}_\lambda(\boldsymbol{D})\boldsymbol{V}^T, \mathcal{S}_\lambda(\boldsymbol{D}) = \text{diag}(\{(d_i-\lambda)_+\}), \tag{15}$$

where $x_+ = max(0,x)$ denotes the non-negative part of x. The following theorem shows that a connection can be established between the SVD-thresholding operation and solving the simplest form of NNR optimization.

Theorem 2 (Proposition 2.1, Chen et al. (2013))**.** *For any $\lambda \geq 0$ and $\boldsymbol{Y} \in \mathbb{R}^{p\times q}$, the hard/soft SVD-thresholding operators can be characterized as*

$$\mathcal{H}_\lambda(\boldsymbol{Y}) = \arg\min_{C} \left\{ \|\boldsymbol{Y}-\boldsymbol{C}\|_F^2 + \lambda^2 r(\boldsymbol{C}) \right\}, \tag{16}$$

$$\mathcal{S}_\lambda(\boldsymbol{Y}) = \arg\min_{C} \left\{ \|\boldsymbol{Y}-\boldsymbol{C}\|_F^2 + \lambda \|\boldsymbol{C}\|_* \right\}, \tag{17}$$

where $r(\boldsymbol{C})$ denotes the rank of matrix $\boldsymbol{C}$.

This theorem suggests that instead of taking the slow gradient descend, one can benefit from

applying the soft SVD-thresholding operator to the observation matrix $\boldsymbol{Y}$ to obtain the exact solution, which involves solving the full SVD of $\boldsymbol{Y}$. Unfortunately this one-step solution can not be directly applied to NNR-GRRLF. This is because: 1) there are two matrices instead of one that are involved in the objective function, with different norm constraints; 2) we have also an additional covariate matrix $\boldsymbol{X}$ and a smoothing matrix $\boldsymbol{H}$ that complicates the objective function.

However, with some reformulations to the problem we can decouple the two matrices involved and apply the above SVD scheme in a step-wise fashion. Using the trick developed in Ji and Ye (2009), we can reduce the convergence rate from Jaggi-Hanzan's $\mathcal{O}(1/k)$ to the optimal rate of $\mathcal{O}(1/k^2)$. But this improved convergence rate does not necessarily imply faster computation in practice because it invokes higher per-iteration cost. We now present the details below.

First consider the minimization of the smooth loss function without the trace norm regularization:

$$\min_{W} f(\boldsymbol{W}). \tag{18}$$

Let $\alpha_k = 1/\beta_k$ be the step size for iteration k, the gradient step for solving the smooth problem

$$\boldsymbol{W}_k = \boldsymbol{W}_{k-1} - \frac{1}{\beta_k}\nabla f(\boldsymbol{W}_{k-1}) \tag{19}$$

can be reformulated equivalently as as a proximal regularization of the linearized function of $f(\boldsymbol{W})$ at $\boldsymbol{W}_{k-1}$ as

$$\boldsymbol{W}_k = \arg\min_{W} P_{\beta_k}(\boldsymbol{W}, \boldsymbol{W}_{k-1}), \tag{20}$$

where

$$P_{\beta_k}(\boldsymbol{W}, \boldsymbol{W}_{k-1}) = f(\boldsymbol{W}_{k-1}) + \langle \boldsymbol{W} - \boldsymbol{W}_{k-1}, \nabla f(\boldsymbol{W}_{k-1})\rangle + \frac{\beta_k}{2}\|\boldsymbol{W} - \boldsymbol{W}_{k-1}\|_F^2,$$

and $\langle \boldsymbol{A}, \boldsymbol{B}\rangle = \mathrm{tr}(\boldsymbol{A}^T\boldsymbol{B})$ denotes the matrix inner product. The above P_{β_k} can be considered as a linear approximation of the function f at point $\boldsymbol{W}_{k-1}$ regularized by a quadratic proximal term.

Based on this equivalence, the optimization problem

$$\min_{W} \left(f(\boldsymbol{W}) + \lambda \|\boldsymbol{W}\|_* \right) \tag{21}$$

for $\lambda \geq 0$ can be solved with the following iterative step:

$$\boldsymbol{W}_k = \arg\min_{W} P_{\beta_k}(\boldsymbol{W}, \boldsymbol{W}_{k-1}) + \lambda \|\boldsymbol{W}\|_* \tag{22}$$

By ignoring $\boldsymbol{W}$-independent terms, we arrive at a new objective

$$Q_{\beta_k}(\boldsymbol{W}, \boldsymbol{W}_{k-1}) = \frac{\beta_k}{2} \left\| \boldsymbol{W} - \left(\boldsymbol{W}_{k-1} - \frac{1}{\beta_k} \nabla f(\boldsymbol{W}_{t-1}) \right) \right\|_F^2 + \lambda \|\boldsymbol{W}\|_*, \tag{23}$$

and

$$\boldsymbol{W}_k = \arg\min_{W} Q_{\beta_k}(\boldsymbol{W}, \boldsymbol{W}_{k-1}). \tag{24}$$

The key idea behind the above formulation is that by exploiting the structure of the trace norm solution (that can be computed exactly), it can be proven the convergence rate of the regularized objective is the same as that of the gradient descend of $f(\boldsymbol{W})$. The Nesterov gradient approach can be further exploited to achieve the optimal convergence rate of $\mathcal{O}(1/k^2)$. Readers are referred to Ji and Ye (2009) for details.

Back to GRRLF, let us denote $G(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{L}}) = \left\| \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{B}}\boldsymbol{H} - \widetilde{\boldsymbol{L}}\boldsymbol{H} \right\|_F^2$, then using the idea above we can decouple the term $\widetilde{\boldsymbol{B}}$ and $\widetilde{\boldsymbol{L}}$ in the objective function as

$$\begin{aligned} \widetilde{P}_{\beta_k}(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{L}}, \widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \;=\; & \frac{\beta_k}{2} \left(\left\| \widetilde{\boldsymbol{B}} - \left(\widetilde{\boldsymbol{B}}_{k-1} - \frac{1}{\beta_k} \nabla_B G(\widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \right) \right\|_F^2 \right. \\ & \left. + \left\| \widetilde{\boldsymbol{L}} - \left(\widetilde{\boldsymbol{L}}_{k-1} - \frac{1}{\beta_k} \nabla_L G(\widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \right) \right\|_F^2 \right), \end{aligned} \tag{25}$$

and define a new surrogate objective function at each iteration as

$$\widetilde{Q}_{\beta_k}(\widetilde{\boldsymbol{B}}, \widetilde{\boldsymbol{L}}, \widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \;=\; \widetilde{P}_{\beta_k} + \lambda_1 \|\widetilde{\boldsymbol{B}}\|_* + \lambda_2 \|\widetilde{\boldsymbol{L}}\|_* = \widetilde{Q}^B_{\beta_k} + \widetilde{Q}^L_{\beta_k}, \tag{26}$$

where

$$\widetilde{Q}^{B}_{\beta_k} = \left\| \widetilde{\boldsymbol{B}} - \left(\widetilde{\boldsymbol{B}}_{k-1} - \frac{1}{\beta_k} \nabla_B G(\widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \right) \right\|_F^2 + \frac{2\lambda_1}{\beta_k} \|\widetilde{\boldsymbol{B}}\|_*, \tag{27}$$

$$\widetilde{Q}^{L}_{\beta_k} = \left\| \widetilde{\boldsymbol{L}} - \left(\widetilde{\boldsymbol{L}}_{k-1} - \frac{1}{\beta_k} \nabla_L G(\widetilde{\boldsymbol{B}}_{k-1}, \widetilde{\boldsymbol{L}}_{k-1}) \right) \right\|_F^2 + \frac{2\lambda_2}{\beta_k} \|\widetilde{\boldsymbol{L}}\|_*. \tag{28}$$

Therefore solving for (26) reduces to the application of soft SVD thresholding to (27) and (28) independently.

The gain in convergence rate is however not for free. In each iteration a full SVD needs to be solved instead of a partial SVD that is required by the JH algorithm. Additionally, we can no longer compute an optimal step size for each iteration as that in JH. To summarize, while we can expect a theoretically optimal solver for NNR-GRRLF, the best implementation is application dependent and relies on careful tuning.

Appendix E GM-GRRLF specifications for the synthetic experiment

While we used the GCV procedure to decide which component is estimated first in the estimation error experiment, we replaced the costly GCV with a simpler heuristic thresholding strategy when computing the empirical p-values in the sensitivity experiment to save time. We first estimate the percentage of variance contributed by the covariates with voxel-wise least squares, if the covariate signal proportion exceeds a specified threshold, the covariate effect will be estimated first in the iterative GM-GRRLF. We set the variance threshold to 20% in this experiment, which gave very similar estimation error distribution compared with that of GCV (not shown). We used the HSIC to test for the association between covariates and estimated latent components, and set the association significance threshold to $p_{\text{thres}} = 10^{-3}$.

Appendix F ADNI study design and subjects

The data used in the preparation of this article were obtained 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 primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials. The Principal Investigator of this initiative is Michael W. Weiner, MD, VA Medical Center and University of California – San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org. 818 subjects were genotyped as part of the ADNI study. However, only 736 unrelated Caucasian subjects identified by self-report and confirmed by MDS analysis (Stein et al., 2010a) were included to reduce population stratification effects. Volumetric brain differences were assessed in 176 AD patients (80 female/96 male; 75.47 ± 7.54 years old), 356 MCI subjects (126 female/230 male; 75.03 ± 7.25 years old), and 204 healthy elderly subjects (94 female/110 male; 76.04 ± 4.98 years old).

Appendix G Data preprocessing

Appendix G.1 MRI images

High-resolution structural brain MRI scans were acquired at 58 ADNI sites with 1.5 T MRI scanners using a sagittal 3D MP-RAGE sequence developed for consistency across sites (Jack et al., 2008) (TR=2400 ms, TE=1000 ms, flip angle=$8\,^{\circ}$, field of view=24 cm, final reconstructed voxel resolution = $0.9375 \times 0.9375 \times 1.2\text{mm}^3$). Images were calibrated with phantom-based geometric corrections to ensure consistency across scanners. Additional image corrections included (Jack et al., 2008): (1) correction of geometric distortions due to gradient nonlinearity, (2) adjustment for image intensity inhomogeneity due to B1 field non-uniformity using calibration scans, (3) reducing residual intensity inhomogeneity, and (4) geometric scaling according to a phantom scan acquired for each subject to adjust for scanner- and session-specific calibration errors. Images were linearly registered with 9 parameters to the International Consortium for Brain Imaging template (ICBM-53) (Mazziotta et al., 2001) to adjust for differences in brain position and scaling.

For TBM analysis, a minimal deformation template was first created for the healthy elderly group to serve as an unbiased average template image to which all other images were warped using a non-linear inverse-consistent elastic intensity-based registration algorithm Leow et al. (2005). Volumetric tissue differences were assessed at each voxel in all individuals by calculating the determinant of the Jacobian matrix of the deformation, which encodes local volume excess or deficit relative to the mean template image. The maps of volumetric tissue differences were then downsampled using trilinear interpolation to $4 \times 4 \times 4\text{mm}^3$ isotropic voxel resolution for computational efficiency. After resampling, $29,479$ voxels remained in the brain mask. *The percentage volumetric difference relative to a population-based brain template at each voxel served as a quantitative measure of brain tissue volume difference for genome-wide association.*

Appendix G.2 Genetic data

Genome-wide genotype data were collected at 620,901 markers on the Human610-Quad BeadChip (Illumina, Inc., San Diego, CA). For details on how genetic data were processed, please see Saykin et al. (2010) and Stein et al. (2010a). Different types of markers were genotyped (including copy number probes), but only SNPs were used in this analysis. Due to the filtering based on Illumina GenCall quality control measures, individual subjects have some residual missing genotypes at random SNPs throughout the dataset. We performed imputation using the software, Mach (version 1.0), to infer the haplotype phase and automatically impute the missing genotype data (Li et al., 2009). The genetic tags are translated into corresponding Reference SNP cluster ID (*rsid*) with a dictionary used in imputation. Chromosome positions of the *rsid*s are mapped according to the GRCh38.p2 reference assembly. We use the gene annotations from Ensembl release 79 (Cunningham et al., 2015), which also mapped to GRCh38.p2 reference assembly, to define the start and end position of the genes. All SNPs fall into the same gene region are considered as belonging to the same gene. We use only the SNPs that have been physically genotyped on the 22 autosomes for the gene grouping and after that, a total of $n_{\text{gene}} = 26,664$ genes were left for analysis. Only SNPs with imputed *minor allele frequency* (MAF) ≥ 0.1 are used for the single-locus experiment on the target gene.

Appendix H Statistical methods for ADNI data analysis

We use a modified version of the LSKM-based vGWAS proposed in Ge et al. (2012) in the ADNI data analysis, which is detailed below.

Appendix H.1 Fitted model and choice of kernel

Since only gender and age are supplied as covariates, the dimension reduction on covariates is unnecessary in this particular case. So we fit the following simplified null model for the GWAS analysis on ADNI data

$$y_{i,\boldsymbol{v}} = \boldsymbol{x}_i^\top \boldsymbol{\beta}_{\boldsymbol{v}} + l_{i,\boldsymbol{v}} + \xi_{i,\boldsymbol{v}},$$

where $i = 1, \cdots, n$ is the subject index, $\boldsymbol{v} \in \Omega$ is the voxel index, y is the image phenotype, $\boldsymbol{x}$ is the covariates, l the latent effect and ξ the residual component. We use a generalized identity by state (IBS) function as the kernel function in this study, which is defined as

$$\kappa(\boldsymbol{g}_1, \boldsymbol{g}_2) = 1 - \|\boldsymbol{g}_1 - \boldsymbol{g}_2\|_1/(2n_{\boldsymbol{g}}),$$

where $\boldsymbol{g}_i \in [0,2]^{n_{\boldsymbol{g}}}$ for $i = 1, 2$ is the genetic data and $n_{\boldsymbol{g}}$ is the number of SNPs on gene $\boldsymbol{g}$. To
805 expedite the computation, we use *incomplete Cholesky decomposition* (ICL) (Bach and Jordan, 2003) to give low rank approximation $LL^\top$ of the kernel matrix K. We restrict the maximum allowed rank to $r = 50$ and the results are similar to those using original kernel matrix (data not shown).

Appendix H.2 Null distribution of the LSKM test score

The test score Q of LSKM follows a mixed chi-square distribution under certain assumptions (see Liu et al. (2007) for details). With the Satterthwaite method (matching the first two moments), the distribution of the test score Q can be approximated by equating the mean and variance of the scaled chi-square variable $\kappa\chi^2_\nu$. Specifically, $\kappa = \widetilde{I}_{\tau\tau}/2\widetilde{e}$, $\nu = 2\widetilde{e}^2/\widetilde{I}_{\tau\tau}$, where $\widetilde{I}_{\tau\tau} = I_{\tau\tau} - I_{\tau\sigma^2} I^{-1}_{\sigma^2\sigma^2} I_{\tau\sigma^2}$, $I_{\tau\tau} = \mathrm{tr}\left((P_0K)^2\right)$, $I_{\tau\sigma^2} = \mathrm{tr}\left(P_0KP_0\right)/2$, $I_{\sigma^2\sigma^2} = \mathrm{tr}\left(P_0^2\right)/2$ and $\widetilde{e} = \mathrm{tr}\left(P_0K\right)/2$, and $\boldsymbol{P}_0$ is the residual forming matrix defined as

$$\boldsymbol{P}_0 = \boldsymbol{I} - \boldsymbol{X}\left(\boldsymbol{X}^\top\boldsymbol{X}\right)^{-1}\boldsymbol{X}^\top.$$

We note however, these assumptions, such as the normality of the residuals, can be easily violated in practice. Also the fitness of the approximation, especially at the tail, depends on how well the moment matched distribution in the scaled chi-square family resembles the originally mixed

chi-square. Thus researchers need to check the validity of the use of parametric approximation with their data. Unfortunately our *null* simulations suggest that with the kernel matrices derived from empirical genetic data, the p-values evaluated using the approximated scaled chi-square is severely inflated at the tail, see Figure S1(a) for the distribution of p-values using $26,664$ empirical kernel matrices from the ADNI1 dataset and *i.i.d* standard Gaussian variables as responses. To correct for the inflation, we use nonparametric permutation to evaluate the p-value under the null instead of the scaled chi-square approximation. The subject index of $\widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}$ is independently shuffled for each voxel $\boldsymbol{v}$, then the test score $Q_{\boldsymbol{v}}^{\text{null}}$ is calculated using the shuffled $\widehat{\boldsymbol{\xi}}_{\boldsymbol{v}}$ for each voxel $\boldsymbol{v}$. $\{Q_{\boldsymbol{v}}^{\text{null}}\}_{\boldsymbol{v}\in\Omega}$ is considered as N_{vox} independent test scores under the null hypothesis thus giving the empirical null distribution. Then it is used to calculate the empirical p-values for the test scores as

$$p_{\text{emp}}(Q) = \max\left\{\frac{\#\{Q_{\boldsymbol{v}}^{\text{null}} \geq Q\}}{N_{\text{vox}}}, \frac{1}{N_{\text{vox}}}\right\}.$$

We further used generalized Pareto distribution (GPD) (Coles et al., 2001; Knijnenburg et al., 2009) to approximate the tail of the empirical distribution. The largest 1% of $\{Q_{\boldsymbol{v}}^{\text{null}}\}$ is used for the maximum likelihood estimation of GPD parameters and then the p-values for the tail statistics are evaluated using the estimated parameters. The results of the GPD approximated p-values are presented in Figure S1(b). The GPD approximated tail p-values are also prone to inflation when they are smaller than 10^{-4}. In this regard, no peak inference is conducted in this study, as the results are unreliable. We report only the result of cluster-size based inference with cluster-forming threshold set to $p_{\text{thres}} = 10^{-3}$, where the inflation is negligible.

Appendix H.3 Cluster size based inference

In this study, the maximum cluster size S in RESEL for each gene is used as the test statistics. RESEL stands for RESolution ELement (Worsley et al., 1992), which represents a virtual voxel with size $[\text{FWHM}_X, \text{FWHM}_Y, \text{FWHM}_Z]$. In the stationary case, RESEL count R is the number

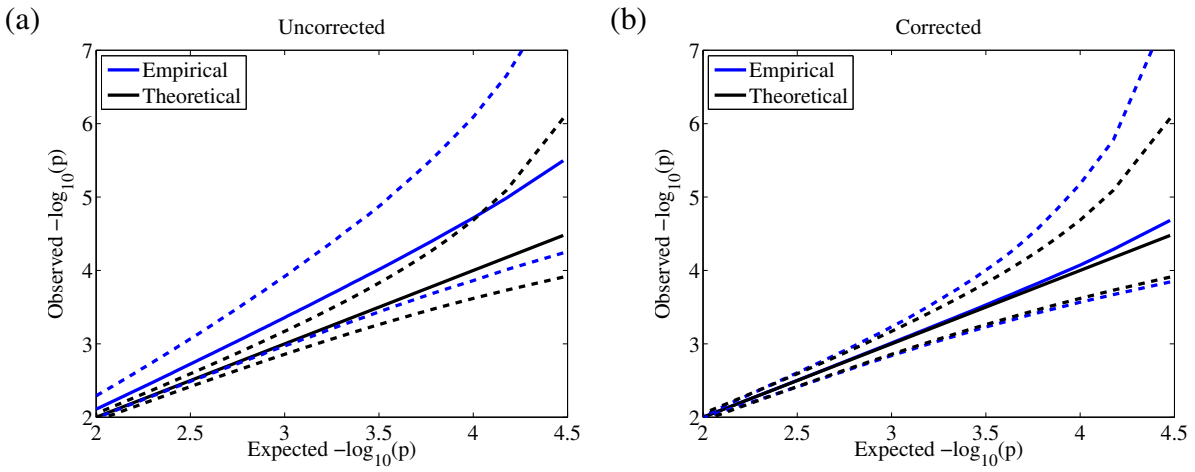


Figure S1: Expected and observed distribution of p-values in $\log_{10}$ scale. (a) Uncorrected p-values. (b) Corrected p-values. Black solid: expected p-value, black dash: expected 95% confidence interval, blue solid: median of the observed p-value, blue dash: observed 95% interval. Uncorrected p-values from the LKSM Satterthwaite approximation gives much more false positives than expected thus can not be directly reported.

of such virtual voxels that fit into the search volume V

$$R = \frac{V}{\prod_{u\in X,Y,Z} \mathrm{FWHM}_u}.$$

In the nonstationary case (Hayasaka et al., 2004), voxel-wise *Resels Per Voxel* (RPV) statistics is defined as

$$\mathrm{RPV}_{\boldsymbol{v}} = \sum_{\boldsymbol{v}\in\Omega} \frac{|\Omega|^{-1}V}{\prod_{u\in X,Y,Z} \mathrm{FWHM}_u^{(\boldsymbol{v})}},$$

where $|\Omega|$ is the voxel count and $R_n = \sum_{\boldsymbol{v}} \mathrm{RPV}_{\boldsymbol{v}}$ generalizes RESEL count R in stationary case. Simply put, RESEL count is a measure of volume normalized by the smoothness of image. Specifically, we use SPM's *spm_est_smoothness* function in SPM 8 to estimate the RPV image. Then we construct all clusters the using *spm_bwlabel* function with the connectivity pattern criterion set to 'edge'. The cluster size is calculated by integrating RPV for each cluster. For each gene, the maximum cluster size is reported. To construct the null distribution of the maximum cluster size, we shuffled the subject index and then permute the rows and columns of the kernel matrices accordingly. For each gene, 20 null statistics were calculated. Then the $M_{\text{perm}} = 20N_{\text{gene}}$ null statistics

were pooled together to give an empirical null distribution $\{S_b^{\text{null}}\}_{b=1}^{M_{\text{perm}}}$. The empirical p-value of the cluster size S is given as

$$p_{\text{emp}}^{\text{clu}}(S) = \max\left\{\frac{\#\{S_b^{\text{null}} \geq S\}}{M_{\text{perm}}}, \frac{1}{M_{\text{perm}}}\right\}.$$

We found that the number of permutations we ran is unable to give sufficient samples for the estimation of tail distribution of maximum cluster size using GPD (data not shown), so only the empirical p-value is reported.

References

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Akaike, H. (1974). A new look at the statistical model identification. *Automatic Control, IEEE Transactions on*, 19(6):716–723.

Avron, H., Kale, S., Kasiviswanathan, S., and Sindhwani, V. (2012). Efficient and practical stochastic subgradient descent for nuclear norm regularization. *arXiv preprint arXiv:1206.6384*.

Bach, F. R. and Jordan, M. I. (2003). Kernel independent component analysis. *The Journal of Machine Learning Research*, 3:1–48.

Ballmaier, M., Toga, A. W., Blanton, R. E., Sowell, E. R., Lavretsky, H., Peterson, J., Pham, D., and Kumar, A. (2014). Anterior cingulate, gyrus rectus, and orbitofrontal abnormalities in elderly depressed patients: an mri-based parcellation of the prefrontal cortex. *American Journal of Psychiatry*.

Batmanghelich, N. K., Dalca, A. V., Sabuncu, M. R., and Golland, P. (2013). Joint modeling of imaging and genetics. In *Information Processing in Medical Imaging: Conference*, pages 766–77.

Bhattacharya, A., Dunson, D. B., et al. (2011). Sparse bayesian infinite factor models. *Biometrika*, 98(2):291.

Bi, H. and Sze, C.-I. (2002). N-methyl-d-aspartate receptor subunit nr2a and nr2b messenger rna levels are altered in the hippocampus and entorhinal cortex in alzheimer's disease. *Journal of the neurological sciences*, 200(1):11–18.

Bigos, K. L., Mattay, V. S., Callicott, J. H., Straub, R. E., Vakkalanka, R., Kolachana, B., Hyde, T. M., Lipska, B. K., Kleinman, J. E., and Weinberger, D. R. (2010). Genetic variation in cacna1c affects brain circuitries related to mental illness. *Archives of General Psychiatry*, 67(9):939–945.

Blackman, A. V., Abrahamsson, T., Costa, R. P., Lalanne, T., and Sjöström, P. J. (2013). Target-cell-specific short-term plasticity in local circuits. *Frontiers in synaptic neuroscience*, 5.

Boumal, N., Mishra, B., Absil, P.-A., and Sepulchre, R. (2014). Manopt, a matlab toolbox for optimization on manifolds. *The Journal of Machine Learning Research*, 15(1):1455–1459.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on*, 56(5):2053–2080.

Chen, K., Dong, H., and Chan, K.-S. (2013). Reduced rank regression via adaptive nuclear norm penalization. *Biometrika*, page ast036.

Cheng, W., Rolls, E., Liu, W., Chang, M., Huang, C.-C., Zhang, J., Xie, P., Lin, C.-P., Wang, F., Qiu, J., and Feng, J. (2015). Medial and lateral orbitofrontal cortex functional connectivity circuit changes in depression. *in preparation*.

Chiang, M.-C., Barysheva, M., McMahon, K. L., de Zubicaray, G. I., Johnson, K., Montgomery, G. W., Martin, N. G., Toga, A. W., Wright, M. J., Shapshak, P., et al. (2012). Gene network effects on brain microstructure and intellectual performance identified in 472 twins. *The Journal of Neuroscience*, 32(25):8732–8745.

Coles, S., Bawa, J., Trenner, L., and Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values*, volume 208. Springer.

Consortium, A.-. et al. (2012). The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in systems neuroscience*, 6.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., et al. (2015). Ensembl 2015. *Nucleic acids research*, 43(D1):D662–D669.

De Leeuw, J. (1994). Block-relaxation algorithms in statistics. In *Information systems and data analysis*, pages 308–324. Springer.

Dolan, J. and Mitchell, K. J. (2013). Mutation of elfn1 in mice causes seizures and hyperactivity. *PloS one*.

Efron, B. (2010). *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*, volume 1. Cambridge University Press.

Erk, S., Meyer-Lindenberg, A., Schmierer, P., Mohnke, S., Grimm, O., Garbusow, M., Haddad, L., Poehland, L., Mühleisen, T. W., Witt, S. H., et al. (2014). Hippocampal and frontolimbic function as intermediate phenotype for psychosis: evidence from healthy relatives and a common risk variant in cacna1c. *Biological psychiatry*, 76(6):466–475.

Eu-ahsunthornwattana, J., Miller, E. N., Fakiola, M., Jeronimo, S. M. B., Blackwell, J. M., Cordell, H. J., and 2, W. T. C. C. C. (2014). Comparison of methods to account for relatedness in genome-wide association studies with family-based data. *PLoS Genet*, 10(7):e1004445.

Farber, N. B., Newcomer, J. W., and Olney, J. W. (1997). The glutamate synapse in neuropsychiatric disorders. focus on schizophrenia and alzheimer's disease. *Progress in brain research*, 116:421–437.

Franke, B., Vasquez, A. A., Veltman, J. A., Brunner, H. G., Rijpkema, M., and Fernández, G. (2010). Genetic variation in cacna1c, a gene associated with bipolar disorder, influences brainstem rather than gray matter volume in healthy individuals. *Biological psychiatry*, 68(6):586–588.

Fusi, N., Stegle, O., and Lawrence, N. D. (2012). Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies. *PLoS Comput Biol*, 8(1):e1002330.

Ganjgahi, H., Winkler, A. M., Glahn, D. C., Blangero, J., Kochunov, P., and Nichols, T. E. (2015). Fast and powerful heritability inference for family-based neuroimaging studies. *NeuroImage*, 115:256–268.

Ge, T., Feng, J., Hibar, D. P., Thompson, P. M., and Nichols, T. E. (2012). Increasing power for voxel-wise genome-wide association studies: the random field theory, least square kernel machines and fast permutation procedures. *Neuroimage*, 63(2):858–873.

Ge, T., Nichols, T. E., Ghosh, D., Mormino, E. C., Smoller, J. W., Sabuncu, M. R., Initiative, A. D. N., et al. (2015a). A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *Neuroimage*, 109:505–514.

Ge, T., Nichols, T. E., Lee, P. H., Holmes, A. J., Roffman, J. L., Buckner, R. L., Sabuncu, M. R., and Smoller, J. W. (2015b). Massively expedited genome-wide heritability analysis (megha). *Proceedings of the National Academy of Sciences*, 112(8):2479–2484.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schölkopf, B., and Smola, A. J. (2007). A kernel statistical test of independence. In *Advances in Neural Information Processing Systems*, volume 20, pages 585–592. MIT Press.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B. (2005). Kernel methods for measuring independence. *The Journal of Machine Learning Research*, 6:2075–2129.

Hardoon, D. R., Ettinger, U., Mourão-Miranda, J., Antonova, E., Collier, D., Kumari, V., Williams, S. C., and Brammer, M. (2009). Correlation-based multivariate analysis of genetic influence on brain volume. *Neuroscience letters*, 450(3):281–286.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., van de Lagemaat, L. N., Smith, K. A., Ebbert, A., Riley, Z. L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, 489(7416):391–399.

Hayasaka, S., Phan, K. L., Liberzon, I., Worsley, K. J., and Nichols, T. E. (2004). Nonstationary cluster-size inference with random field and permutation methods. *Neuroimage*, 22(2):676–687.

Hazan, E. (2008). Sparse approximate solutions to semidefinite programs. In *LATIN 2008: Theoretical Informatics*, pages 306–316. Springer.

Heatherton, T. (1991). The fagerstrom test for nicotine dependence, a revision of the fagerstrom tolerance questionnarire. *Br J Addict*, 86(9):1119–1127.

Hibar, D. P., Stein, J. L., Kohannim, O., Jahanshad, N., Saykin, A. J., Shen, L., Kim, S., Pankratz, N., Foroud, T., Huentelman, M. J., et al. (2011). Voxelwise gene-wide association study (vgenewas): multivariate gene-based association testing in 731 elderly subjects. *Neuroimage*, 56(4):1875–1891.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N., Toro, R., Wittfeld, K., Abramovic, L., Andersson, M., et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature*, 520(7546):224–229.

Hopp, S., DAngelo, H., Royer, S., Kaercher, R., Adzovic, L., and Wenk, G. (2014). Differential rescue of spatial memory deficits in aged rats by l-type voltage-dependent calcium channel and ryanodine receptor antagonism. *Neuroscience*, 280:10–18.

Hruz, T., Laule, O., Szabo, G., Wessendorp, F., Bleuler, S., Oertle, L., Widmayer, P., Gruissem, W., and Zimmermann, P. (2008). Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Advances in bioinformatics*, 2008.

Hsieh, C.-J. and Olsen, P. (2014). Nuclear norm minimization via active subspace selection. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 575–583.

Hua, W.-Y. and Ghosh, D. (2014). Equivalence of kernel machine regression and kernel distance covariance for multidimensional trait association studies. *arXiv preprint arXiv:1402.2679*.

Hua, W.-Y., Nichols, T. E., Ghosh, D., Initiative, A. D. N., et al. (2015). Multiple comparison procedures for neuroimaging genomewide association studies. *Biostatistics*, 16(1):17–30.

Huang, M., Nichols, T., Huang, C., Yu, Y., Lu, Z., Knickmeyer, R. C., Feng, Q., and Zhu, H. (2015). Fvgwas: Fast voxelwise genome wide association analysis of large-scale imaging genetic data. *NeuroImage*, 118:613 – 627.

Hynd, M. R., Scott, H. L., and Dodd, P. R. (2004). Differential expression of n-methyl-d-aspartate receptor nr2 isoforms in alzheimer's disease. *Journal of neurochemistry*, 90(4):913–919.

Hyvärinen, A., Karhunen, J., and Oja, E. (2004). *Independent component analysis*, volume 46. John Wiley & Sons.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of multivariate analysis*, 5(2):248–264.

Jack, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging*, 27(4):685–691.

Jaggi, M., Sulovsk, M., et al. (2010). A simple algorithm for nuclear norm regularized problems. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pages 471–478.

Ji, S. and Ye, J. (2009). An accelerated gradient method for trace norm minimization. In *Proceedings of the 26th annual international conference on machine learning*, pages 457–464. ACM.

Jia, T., Macare, C., Desrivières, S., Gonzalez, D. A., Tao, C., Ji, X., Ruggeri, B., Nees, F., Banaschewski, T., Barker, G. J., et al. (2016). Neural basis of reward anticipation and its genetic determinants. *Proceedings of the National Academy of Sciences*, page 201503252.

Jiang, B. and Liu, J. S. (2015). Bayesian partition models for identifying expression quantitative trait loci. *Journal of the American Statistical Association*, 110(512):1350–1361.

Jiang, H. and Jia, J. (2009). Association between nr2b subunit gene (grin2b) promoter polymorphisms and sporadic alzheimers disease in the north chinese population. *Neuroscience letters*, 450(3):356–360.

Joyner, A. H., Bloss, C. S., Bakken, T. E., Rimol, L. M., Melle, I., Agartz, I., Djurovic, S., Topol, E. J., Schork, N. J., Andreassen, O. A., et al. (2009). A common mecp2 haplotype associates with reduced cortical surface area in humans in two independent populations. *Proceedings of the National Academy of Sciences*, 106(36):15483–15488.

Karasuyama, M. and Sugiyama, M. (2012). Canonical dependency analysis based on squared-loss mutual information. *Neural networks*, 34:46–55.

Knijnenburg, T. A., Wessels, L. F., Reinders, M. J., and Shmulevich, I. (2009). Fewer permutations, more accurate p-values. *Bioinformatics*, 25(12):i161–i168.

Koran, M. E. I., Hohman, T. J., and Thornton-Wells, T. A. (2014). Genetic interactions found between calcium channel genes modulate amyloid load measured by positron emission tomography. *Human genetics*, 133(1):85–93.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, (8):30–37.

Lamprecht, R. and LeDoux, J. (2004). Structural plasticity and memory. *Nature Reviews Neuroscience*, 5(1):45–54.

Lange, K. (2010). *Numerical analysis for statisticians*. Springer Science & Business Media.

Laurent, M. and Vallentin, F. (2012). *Semidefinite Optimization*.

Le Floch, É., Guillemot, V., Frouin, V., Pinel, P., Lalanne, C., Trinchera, L., Tenenhaus, A., Moreno, A., Zilbovicius, M., Bourgeron, T., et al. (2012). Significant correlation between a set of genetic polymorphisms and a functional brain network revealed by feature selection and sparse partial least squares. *Neuroimage*, 63(1):11–24.

Le Floch, E., Trinchera, L., Guillemot, V., Tenenhaus, A., Poline, J.-B., Frouin, V., and Duchesnay, E. (2013). Dimension reduction and regularization combined with partial least squares in high dimensional imaging genetics studies. In *New Perspectives in Partial Least Squares and Related Methods*, pages 147–158. Springer.

Lee, T.-L., Raygada, M. J., and Rennert, O. M. (2012). Integrative gene network analysis provides novel regulatory relationships, genetic contributions and susceptible targets in autism spectrum disorders. *Gene*, 496(2):88–96.

Leow, A., Huang, S.-C., Geng, A., Becker, J., Davis, S., Toga, A., and Thompson, P. (2005). Inverse consistent mapping in 3d deformable image registration: its construction and statistical properties. In *Information Processing in Medical Imaging*, pages 493–503. Springer.

Li, M. D., Burns, T. C., Morgan, A. A., and Khatri, P. (2014). Integrated multi-cohort transcriptional meta-analysis of neurodegenerative diseases. *Acta Neuropathol Commun*, 2:93.

Li, X. (2014). *Tensor Based Statistical Models with Applications in Neuroimaging Data Analysis.* PhD thesis, North Carolina State University.

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype imputation. *Annual review of genomics and human genetics*, 10:387.

Liang, L. and Wei, H. (2015). Dantrolene, a treatment for alzheimer disease? *Alzheimer Disease & Associated Disorders*, 29(1):1–5.

Lin, D., Li, J., Calhoun, V. D., and Wang, Y.-P. (2015). Detection of genetic factors associated with multiple correlated imaging phenotypes by a sparse regression model. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 1368–1371. IEEE.

Liu, D., Lin, X., and Ghosh, D. (2007). Semiparametric regression of multidimensional genetic pathway data: Least-squares kernel machines and linear mixed models. *Biometrics*, 63(4):1079–1088.

Liu, J. and Calhoun, V. D. (2014). A review of multivariate analyses in imaging genetics. *Frontiers in neuroinformatics*, 8.

Liu, J., Pearlson, G., Windemuth, A., Ruano, G., Perrone-Bizzozero, N. I., and Calhoun, V. (2009). Combining fmri and snp data to investigate connections between brain function and genetics using parallel ica. *Human brain mapping*, 30(1):241–255.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., et al. (2001). A probabilistic atlas and reference system for the human brain: International consortium for brain mapping (icbm). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1412):1293–1322.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, volume 37. CRC press.

Mishra, B., Meyer, G., Bach, F., and Sepulchre, R. (2013). Low-rank optimization with trace norm penalty. *SIAM Journal on Optimization*, 23(4):2124–2149.

Montagna, S., Tokdar, S. T., Neelon, B., and Dunson, D. B. (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68(4):1064–1073.

Nebion, A. (2014). Genevisible. http://genevisible.com/. Accessed: 2015-09-28.

Noebels, J. L., Avoli, M., Rogawski, M. A., Olsen, R. W., Delgado-Escueta, A. V., Grisar, T., Lakaye, B., de Nijs, L., LoTurco, J., Daga, A., et al. (2012). Myoclonin1/efhc1 in cell division, neuroblast migration, synapse/dendrite formation in juvenile myoclonic epilepsy. In *Jasper's Basic Mechanisms of the Epilepsies [Internet]. 4th edition*. National Center for Biotechnology Information (US).

Ojelade, S. A., Jia, T., Rodan, A. R., Chenyang, T., Kadrmas, J. L., Cattrell, A., Ruggeri, B., Charoen, P., Lemaitre, H., Banaschewski, T., et al. (2015). Rsu1 regulates ethanol consumption in drosophila and humans. *Proceedings of the National Academy of Sciences*, 112(30):E4085–E4093.

Oliva, C. A., Vargas, J. Y., and Inestrosa, N. C. (2013). Wnts in adult brain: from synaptic plasticity to cognitive deficiencies. *Frontiers in cellular neuroscience*, 7.

Öngür, D., Ferry, A. T., and Price, J. L. (2003). Architectonic subdivision of the human orbital and medial prefrontal cortex. *Journal of Comparative Neurology*, 460(3):425–449.

Parsons, C. G., Stöffler, A., and Danysz, W. (2007). Memantine: a nmda receptor antagonist that improves memory by restoration of homeostasis in the glutamatergic system-too little activation is bad, too much is even worse. *Neuropharmacology*, 53(6):699–723.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., and Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images: the analysis of functional brain images*. Academic press.

Petryshen, T. L., Sabeti, P. C., Aldinger, K. A., Fry, B., Fan, J. B., Schaffner, S., Waggoner, S. G., Tahl, A. R., and Sklar, P. (2010). Population genetic study of the brain-derived neurotrophic factor (bdnf) gene. *Molecular psychiatry*, 15(8):810–815.

PGC et al. (2013). Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *The Lancet*, 381(9875):1371–1379.

Poline, J.-B., Breeze, J., and Frouin, V. (2015). Imaging genetics with fmri. In *fMRI: From Nuclear Spins to Brain Functions*, pages 699–738. Springer.

Potkin, S. G., Turner, J. A., Guffanti, G., Lakatos, A., Fallon, J. H., Nguyen, D. D., Mathalon, D., Ford, J., Lauriello, J., Macciardi, F., et al. (2009). A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia Bulletin*, 35(1):96–108.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition.

Reiss, P. T. and Ogden, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics*, 66(1):61–69.

Richiardi, J., Altmann, A., Milazzo, A.-C., Chang, C., Chakravarty, M. M., Banaschewski, T., Barker, G. J., Bokde, A. L., Bromberg, U., Büchel, C., et al. (2015). Correlated gene expression supports synchronous activity in brain networks. *Science*, 348(6240):1241–1244.

Riise, J., Plath, N., Pakkenberg, B., and Parachikova, A. (2015). Aberrant wnt signaling pathway in medial temporal lobe structures of alzheimers disease. *Journal of Neural Transmission*, pages 1–16.

Saunders, J. B., Aasland, O. G., Babor, T. F., Fuente, J. R. D. L., and Grant, M. (1993). Development of the alcohol use disorders identification test (audit): Who collaborative project on early detection of persons with harmful alcohol consumption-ii. *Addiction*, 88(6):791–804.

Saykin, A. J., Shen, L., Foroud, T. M., Potkin, S. G., Swaminathan, S., Kim, S., Risacher, S. L., Nho, K., Huentelman, M. J., Craig, D. W., et al. (2010). Alzheimer's disease neuroimaging initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans. *Alzheimer's & Dementia*, 6(3):265–273.

Schwab, K. R., Patterson, L. T., Hartman, H. A., Song, N., Lang, R. A., Lin, X., and Potter, S. S. (2007). Pygo1 and pygo2 roles in wnt signaling in mammalian kidney development. *BMC biology*, 5(1):15.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464.

Shinawi, M., Sahoo, T., Maranda, B., Skinner, S., Skinner, C., Chinault, C., Zascavage, R., Peters, S. U., Patel, A., Stevenson, R. E., et al. (2011). 11p14. 1 microdeletions associated with adhd, autism, developmental delay, and obesity. *American Journal of Medical Genetics Part A*, 155(6):1272–1280.

Smith, S. M., Miller, K. L., Moeller, S., Xu, J., Auerbach, E. J., Woolrich, M. W., Beckmann, C. F., Jenkinson, M., Andersson, J., Glasser, M. F., et al. (2012). Temporally-independent functional modes of spontaneous brain activity. *Proceedings of the National Academy of Sciences*, 109(8):3131–3136.

Stegle, O., Parts, L., Durbin, R., and Winn, J. (2010). A bayesian framework to account for complex non-genetic factors in gene expression levels greatly increases power in eqtl studies. *PLoS Comput Biol*, 6(5):e1000770.

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W., Saykin, A. J., Shen, L., Foroud, T., Pankratz, N., et al. (2010a). Voxelwise genome-wide association study (vgwas). *Neuroimage*, 53(3):1160–1174.

Stein, J. L., Hua, X., Morra, J. H., Lee, S., Hibar, D. P., Ho, A. J., Leow, A. D., Toga, A. W., Sul, J. H., Kang, H. M., et al. (2010b). Genome-wide analysis reveals novel genes influencing temporal lobe structure with relevance to neurodegeneration in alzheimer's disease. *Neuroimage*, 51(2):542–554.

Stingo, F. C., Guindani, M., Vannucci, M., and Calhoun, V. D. (2013). An integrative bayesian modeling approach to imaging genetics. *Journal of the American Statistical Association*, 108(503):876–891.

Stogmann, E., Lichtner, P., Baumgartner, C., Bonelli, S., Assem-Hilger, E., Leutmezer, F., Schmied, M., Hotzy, C., Strom, T., Meitinger, T., et al. (2006). Idiopathic generalized epilepsy phenotypes associated with different efhc1 mutations. *Neurology*, 67(11):2029–2031.

Sultana, R., Boyd-Kimball, D., Cai, J., Pierce, W. M., Klein, J. B., Merchant, M., and Butterfield, D. A. (2007). Proteomics analysis of the alzheimer's disease hippocampal proteome. *Journal of Alzheimer's disease: JAD*, 11(2):153–164.

Suzuki, T., Delgado-Escueta, A. V., Aguan, K., Alonso, M. E., Shi, J., Hara, Y., Nishida, M., Numata, T., Medina, M. T., Takeuchi, T., et al. (2004). Mutations in efhc1 cause juvenile myoclonic epilepsy. *Nature genetics*, 36(8):842–849.

Suzuki, T. and Sugiyama, M. (2013). Sufficient dimension reduction via squared-loss mutual information estimation. *Neural computation*, 25(3):725–758.

Sylwestrak, E. L. and Ghosh, A. (2012). Elfn1 regulates target-specific release probability at ca1-interneuron synapses. *Science*, 338(6106):536–540.

Tang, Y.-P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., Liu, G., and Tsien, J. Z. (1999). Genetic enhancement of learning and memory in mice. *Nature*, 401(6748):63–69.

Tesli, M., Skatun, K. C., Ousdal, O. T., Brown, A. A., Thoresen, C., Agartz, I., Melle, I., Djurovic, S., Jensen, J., and Andreassen, O. A. (2013). Cacna1c risk variant and amygdala activity in bipolar disorder, schizophrenia and healthy controls. *PloS one*, 8(2):e56970.

Thompson, P. M., Ge, T., Glahn, D. C., Jahanshad, N., and Nichols, T. E. (2013). Genetics of the connectome. *Neuroimage*, 80:475–488.

Thompson, P. M., Stein, J. L., Medland, S. E., Hibar, D. P., Vasquez, A. A., Renteria, M. E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The enigma consortium: large-scale collaborative analyses of neuroimaging and genetic data. *Brain imaging and behavior*, 8(2):153–182.

Tomioka, N. H., Yasuda, H., Miyamoto, H., Hatayama, M., Morimura, N., Matsumoto, Y., Suzuki, T., Odagawa, M., Odaka, Y. S., Iwayama, Y., et al. (2014). Elfn1 recruits presynaptic mglur7 in trans and its loss results in seizures. *Nature communications*, 5.

Van De Ville, D., Seghier, M. L., Lazeyras, F., Blu, T., and Unser, M. (2007). Wspm: Wavelet-based statistical parametric mapping. *Neuroimage*, 37(4):1205–1217.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., Consortium, W.-M. H., et al. (2013). The wu-minn human connectome project: an overview. *Neuroimage*, 80:62–79.

Vounou, M., Janousova, E., Wolz, R., Stein, J. L., Thompson, P. M., Rueckert, D., Montana, G., Initiative, A. D. N., et al. (2012). Sparse reduced-rank regression detects genetic associations with voxel-wise longitudinal phenotypes in alzheimer's disease. *Neuroimage*, 60(1):700–716.

Vounou, M., Nichols, T. E., Montana, G., Initiative, A. D. N., et al. (2010). Discovering genetic associations with high-dimensional neuroimaging phenotypes: a sparse reduced-rank regression approach. *Neuroimage*, 53(3):1147–1159.

Wahba, G. (1990). *Spline models for observational data*, volume 59. Siam.

Wang, H., Nie, F., Huang, H., Kim, S., Nho, K., Risacher, S. L., Saykin, A. J., Shen, L., et al. (2012a). Identifying quantitative trait loci via group-sparse multitask regression and feature selection: an imaging genetics study of the adni cohort. *Bioinformatics*, 28(2):229–237.

Wang, H., Nie, F., Huang, H., Risacher, S. L., Saykin, A. J., Shen, L., et al. (2012b). Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics*, 28(12):i127–i136.

Wang, X., Nan, B., Zhu, J., Koeppe, R., et al. (2014). Regularized 3d functional regression for brain image data via haar wavelets. *The Annals of Applied Statistics*, 8(2):1045–1064.

Woicik, P. A., Stewart, S. H., Pihl, R. O., and Conrod, P. J. (2009). The substance use risk profile scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviors*, 34(12):1042–1055.

Worsley, K. J., Evans, A. C., Marrett, S., and Neelin, P. (1992). A three-dimensional statistical analysis for cbf activation studies in human brain. *Journal of Cerebral Blood Flow & Metabolism*, 12(6):900–918.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., Evans, A. C., et al. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human brain mapping*, 4(1):58–73.

Yang, T., Wang, J., Sun, Q., Hibar, D., Jahanshad, N., Liu, L., Wang, Y., Zhan, L., Thompson, P., and Ye, J. (2015). Detecting genetic risk factors for alzheimer's disease in whole genome sequence data via lasso screening. In *Biomedical Imaging (ISBI), 2015 IEEE 12th International Symposium on*, pages 985–989.

Yashiro, K. and Philpot, B. D. (2008). Regulation of nmda receptor subunit expression and its implications for ltd, ltp, and metaplasticity. *Neuropharmacology*, 55(7):1081–1094.

Yoshimizu, T., Pan, J., Mungenast, A., Madison, J., Su, S., Ketterman, J., Ongur, D., McPhie, D., Cohen, B., Perlis, R., et al. (2014). Functional implications of a psychiatric risk variant within cacna1c in induced human neurons. *Molecular psychiatry*.

Yuan, M., Ekici, A., Lu, Z., and Monteiro, R. (2007). Dimension reduction and coefficient estimation in multivariate linear regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(3):329–346.

Zhang, Q., Shen, Q., Xu, Z., Chen, M., Cheng, L., Zhai, J., Gu, H., Bao, X., Chen, X., Wang, K., et al. (2012a). The effects of cacna1c gene polymorphism on spatial working memory in both healthy controls and patients with schizophrenia or bipolar disorder. *Neuropsychopharmacology*, 37(3):677–684.

Zhang, X., Schuurmans, D., and Yu, Y.-l. (2012b). Accelerated training for matrix-norm regularization: A boosting approach. In *Advances in Neural Information Processing Systems*, pages 2906–2914.

Zhang, Y. and Liu, J. S. (2007). Bayesian inference of epistatic interactions in case-control studies. *Nature genetics*, 39(9):1167–1173.

Zhou, H., Li, L., and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502):540–552.

Zhou, K., Yang, Y., Gao, L., He, G., Li, W., Tang, K., Ji, B., Zhang, M., Li, Y., Yang, J., et al. (2010). Nmda receptor hypofunction induces dysfunctions of energy metabolism and semaphorin signaling in rats: a synaptic proteome study. *Schizophrenia bulletin*, page sbq132.

Zhu, H., Khondker, Z., Lu, Z., and Ibrahim, J. G. (2014). Bayesian generalized low rank regression models for neuroimaging phenotypes and genetic markers. *Journal of the American Statistical Association*, 109(507):977–990.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.