CAPTURING SHARED AND INDIVIDUAL INFORMATION IN FMRI DATA

Javier S. Turek[⋆]*, Cameron T. Ellis*[†]*, Lena J. Skalaban*[†]*, Nicholas B. Turk-Browne*[†]*, Theodore L. Willke*[⋆]

[⋆]Parallel Computing Lab, Intel Labs, Hillsboro, OR 97124
[†]Department of Psychology, Yale University, New Haven, CT 06511

ABSTRACT

Cognitive neuroscience seeks to explain the organization of the brain, but typically focuses on aspects that are shared across people rather than those that vary across individuals. Here, we present a new method for analyzing brain imaging data that captures both shared and individual components of brain activity. Inspired by the shared response model (SRM) and the robust principal components analysis, the robust shared response model (RSRM) aligns functional topographies across humans while preserving a component of sparse, individual activity. Experimental results on adult data showed that RSRM performs as well as or better than SRM, while at the same time capturing reliable markers of individual variability. In a test case of participants with extreme variability, we found that RSRM was able to improve the accuracy more than 60% over SRM for the coding of infant fMRI data.

Index Terms— functional alignment, factor analysis, functional magnetic resonance imaging, individual variability

1. INTRODUCTION

The goal of cognitive neuroscience has predominantly been to characterize brain function at a population level. That is, to draw inferences about how a stranger's brain should work based on a sample of other individuals. To draw such inferences, brain imaging data (typically from functional magnetic resonance imaging, fMRI) are averaged across the sample. This is traditionally done with a general linear model that treats individual variation as a random effect. More recent approaches, such as the shared response model (SRM) [1] and others [2, 3, 4, 5, 6], characterize population-level activity by aligning functional topographies and identifying shared components in the temporal dynamics of activity across the brain to a common stimulus, such as a movie or story.

Although such approaches are incredibly useful, there are also situations where the individual variation they obscure is the target of study. For example, in a clinical context, each patient has a distinct life history, genetic makeup, and constellation of illnesses or disorders. Thus, individual variation may be crucial when using brain imaging for personalized treatment or diagnosis. Even in the healthy brain, individuals vary considerably in different behavioral abilities, from memory capacity, to reading, and to math problem solving.

Development is another domain where such variation is salient. Over the course of development, variation in functional brain activity could arise both within the same individual at different ages (as they develop new abilities) and across individuals at a given age (because their developmental trajectories are misaligned). At the same time, there are also shared components within individuals over development (because past stages serve as the baseline for future stages) and across individuals of a specific age (what tends to happen at that age).

Here, we describe a new method for extracting such individual variation. The method separates shared and individual components of brain activity. Unlike standard functional alignment methods, the individual components are distinguished from noise, which is estimated separately. Specifically, we combine SRM with the idea behind robust principal component analysis [7, 8], and thus we refer to it as the robust shared response model (RSRM). In what follows, we first define the model and derive an algorithm to estimate it. Next, we test its ability to recover signals on synthetic data. Then, we evaluate its performance with both adult and (rare) infant fMRI data. RSRM exhibits higher gains, in particular, when analyzing infant data that have greater individual variation.

2. BACKGROUND

Let $\left\{\mathbf{X}^{(i)}\right\}_{i=1}^{N}$ be a set of N fMRI scans, where $\mathbf{X}^{(i)}$ is the measured brain activity for subject i. The data matrix $\mathbf{X}^{(i)} \in \mathbb{R}^{v_i \times t}$ is comprised of t vectorized volumes (TR) in columns of v_i voxels each. The Shared Response Model (SRM) [1] suggests that there is a k-dimensional shared and latent subspace across subjects that can represent data as $\mathbf{X}^{(i)} = \mathbf{W}^{(i)}\mathbf{R} + \mathbf{E}^{(i)}$ for $i = 1 \dots N$, where $\mathbf{R} \in \mathbb{R}^{k \times t}$ is the k-dimensional shared response, $\mathbf{W}^{(i)} \in \mathbb{R}^{v_i \times k}$ is the subject i mapping from shared space to a subject's voxel space, and $\mathbf{E}^{(i)} \in \mathbb{R}^{v_i \times t}$ is an additive noise term for the respective subject. The model assumes the mappings $\mathbf{W}^{(i)}$ to be in the Stiefel manifold $\mathcal{V}_{v_i,k}$ domain, i.e., orthogonal ${\mathbf{W}^{(i)}}^T\mathbf{W}^{(i)} = \mathbf{I}$. In [1], the authors propose a deterministic and a probabilistic version of the model, and approximate them using a block-coordinate descent approach [9] and an expectation-maximization scheme [10], respectively.

This model has been shown to have several benefits and practical extensions: addition of new data and subjects [1], increased prediction accuracy [1, 11], computational advantages

[12], and semi-supervision [13]. SRM has been designed to look for shared representations between people; however, if the participants being studied have substantial differences then this method may not be optimal.

3. THE ROBUST SHARED RESPONSE MODEL

To improve SRM, we have modified the model to capture individual information. This robust shared response model (RSRM) introduces a new component that captures outlying information. This component reflects variability in individual participants that is not accounted for by the shared representation. In RSRM, each subject's data is described by a summation of the shared response, this individual term, and a residual noise component as follows:

$$\mathbf{X}^{(i)} = \mathbf{W}^{(i)}\mathbf{R} + \mathbf{S}^{(i)} + \mathbf{E}^{(i)} \qquad \forall i = 1 \dots N, \quad (1)$$

where $\mathbf{S}^{(i)} \in \mathbb{R}^{v_i \times t}$ is the individual component for subject i. Similar to SRM, we constrain the mappings $\mathbf{W}^{(i)}$ of RSRM to the Stiefel manifold $\mathcal{V}_{v_i,k}$. Ideally, the $\mathbf{S}^{(i)}$ matrix should contain structured information that is not represented by the shared component. We further assume that individual data is infrequent, so these matrices are sparse. In that sense, RSRM follows the robust principal component analysis and related approaches [7, 8]. When stacking the subjects, RSRM resembles these models. However, the additional constraints requiring orthogonality of each mapping $\mathbf{W}^{(i)}$, makes RSRM a model that yields a different solution.

3.1. Estimating the model

We propose to estimate the $\mathbf{W}^{(i)}$, $\mathbf{S}^{(i)}$, and $\mathbf{R}$ components in Eq. (1) by solving an optimization problem. We use ℓ_1-regularization to obtain sparse $\mathbf{S}^{(i)}$ components:

$$\min_{\mathbf{S}^{(i)},\mathbf{W}^{(i)},\mathbf{R}} \sum_{i=1}^{N} \tfrac{1}{2}\|\mathbf{X}^{(i)} - \mathbf{W}^{(i)}\mathbf{R} - \mathbf{S}^{(i)}\|_F^2 + \lambda_i\|\mathbf{S}^{(i)}\|_1 \quad (2)$$

$$\text{s.t.} \qquad {\mathbf{W}^{(i)}}^T\mathbf{W}^{(i)} = \mathbf{I} \qquad \forall i = 1 \dots N,$$

where $\lambda_i > 0$ is a regularization parameter that balances between the fidelity to the data and the sparsity of $\mathbf{S}^{(i)}$ for each subject. The ℓ_1-norm term is defined as $\|A\|_1 = \sum_{ij}|A_{ij}|$. The value of the regularization parameter determines the amount of data that is considered individual and shared. Thus, it is worth noting that when $\lambda_i = \lambda$ for all i, and $\lambda \to \infty$, Problem (2) is equivalent to the deterministic problem of SRM, as the sparse matrices $\mathbf{S}^{(i)} \to \mathbf{0}$. On the other hand, when $\lambda \to 0$ the optimal solution to the problem is given by $\mathbf{S}^{(i)} \to \mathbf{X}^{(i)}$ and $\mathbf{R} \to \mathbf{0}$.

3.2. Computing the solution

The optimization task in Problem (2) is non-convex. Therefore, we approximate a solution by applying the Block Coordinate Descent approach [9]. This approach partitions the variables into blocks, and iteratively optimizes each block while fixing the values of the variables in other blocks. In the RSRM case, we define a block for each of the matrices $\mathbf{W}^{(i)}$, $\mathbf{S}^{(i)}$, and $\mathbf{R}$. Despite the non-convexity of Problem (2), the subproblems for each variable are convex with well defined solutions.

The shared response term is computed by fixing the values of all variables while minimizing $\mathbf{R}$. The constrained subproblem that depends on $\mathbf{R}$, aims to minimize the sum of Frobenius norm terms alone. This problem has a closed-form solution, and an update to the shared response is given by

$$\mathbf{R} = \frac{1}{N}\sum_{i=1}^{N} {\mathbf{W}^{(i)}}^T \left(\mathbf{X}^{(i)} - \mathbf{S}^{(i)}\right). \quad (3)$$

Intuitively, Eq. (3) suggests that the updated shared response is obtained as the average of all the subjects' projected data after the individual data are removed.

Computing each of the individual terms $\mathbf{S}^{(i)}$ requires the solution of an ℓ_1-regularized minimization task. Each subject's individual term $\mathbf{S}^{(i)}$ in the summation in Problem (2) is decoupled from the others. Therefore, the subproblem to solve for subject i has the form $\min_{\mathbf{S}^{(i)}} \frac{1}{2}\|\mathbf{D}^{(i)} - \mathbf{S}^{(i)}\|_F^2 + \lambda\|\mathbf{S}^{(i)}\|_1$, where $\mathbf{D}^{(i)} = \mathbf{X}^{(i)} - \mathbf{W}^{(i)}\mathbf{R}$ is the individual residual. All the elements in the matrix $\mathbf{S}^{(i)}$ are decoupled, yielding a one-dimensional problem for each element of $\mathbf{S}^{(i)}$. Its solution is obtained by applying the soft-thresholding [14] operator $\mathcal{S}_{\lambda_i}(d)$ on each element d of the matrix $\mathbf{D}^{(i)}$, where the operator is defined as

$$s = \mathcal{S}_{\lambda_i}(d) = \begin{cases} (|d| - \lambda_i)\,\text{sign}(d), & \text{if } |d| > \lambda_i \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Therefore, a single application of the soft-shrinkage operator solves the convex optimization problem for each $\mathbf{S}^{(i)}$.

Next, we update the mappings $\mathbf{W}^{(i)}$ for each subject. The mappings $\mathbf{W}^{(i)}$ are also decoupled in the summation in Problem (2), and thus, can be computed independently per subject. Fixing the shared response, and individual terms yields an optimization problem with penalty function $\|\mathbf{X}^{(i)} - \mathbf{S}^{(i)} - \mathbf{W}^{(i)}\mathbf{R}\|_F^2$ on the Steifel manifold domain for $\mathbf{W}^{(i)}$. Such a minimization task is known as a Procrustes [15] problem and has the closed-form solution

$$\mathbf{W}^{(i)} \quad = \quad \mathbf{U}^{(i)}{\mathbf{V}^{(i)}}^T, \quad (5)$$

where $\mathbf{U}^{(i)}$ and $\mathbf{V}^{(i)}$ are the left and right singular vectors of the matrix $\left(\mathbf{X}^{(i)} - \mathbf{S}^{(i)}\right)\mathbf{R}^T$.

The block coordinate descent approach considered here computes a solution to Problem (2) applying Eq. (3), (4), and (5) iteratively. The method initializes the mappings with a random orthogonal matrix, and the individual terms to the zero matrix. With these two terms set, the shared response can be updated first without the need for an initial value. The runtime complexity of the method is the aggregation of each block update. Governed by matrix multiplies, the updates of $\mathbf{R}$ and all $\mathbf{S}^{(i)}$ have complexity $\mathcal{O}(Nkv_it)$. The mappings $\mathbf{W}^{(i)}$ updates require additional matrix multiplies and SVD computations. The former has a complexity $\mathcal{O}(Nkv_it)$ and the latter $\mathcal{O}\left(Nv_ik^2\right)$. Therefore, the overall runtime complexity of RSRM is $\mathcal{O}(Nkv_it)$, which is the same of SRM.

3.3. Adding new data

RSRM is designed to deal with individual information better than SRM, while preserving many of its properties. A denoised version of the data can be obtained from the shared component $\mathbf{W}^{(i)}\mathbf{R}$. RSRM can generalize to new data from the same subject, or even to new subjects. Assuming that the model generalizes to new data $\tilde{\mathbf{X}}^{(i)}$ for the same subjects, the mappings $\mathbf{W}^{(i)}$ should be the same. To compute the shared response $\tilde{\mathbf{R}}$ and the individual terms $\tilde{\mathbf{S}}^{(i)}$ for the new data, we need to minimize $\|\tilde{\mathbf{X}}^{(i)} - \mathbf{W}^{(i)}\tilde{\mathbf{R}} - \tilde{\mathbf{S}}^{(i)}\|_F^2 + \lambda_i\|\tilde{\mathbf{S}}^{(i)}\|_1$, and obtain $\tilde{\mathbf{S}}^{(i)}$ and $\tilde{\mathbf{R}}$. A solution is computed by iteratively solving Eq. (3) and (4) with the new data.

Alternatively, we can introduce a new subject to the model. For this purpose we need to estimate a mapping $\mathbf{W}^{(new)}$ for the new subject, and also the respective individual term $\mathbf{S}^{(new)}$. We use the shared response $\mathbf{R}$ as a template and minimize the function $\|\mathbf{X}^{(new)} - \mathbf{S}^{(new)} - \mathbf{W}^{(new)}\mathbf{R}\|_F^2 + \lambda_{new}\|\mathbf{S}^{(new)}\|_1$. The terms $\mathbf{W}^{(new)}$ and $\mathbf{S}^{(new)}$ are computed solving iteratively for each matrix with Eq. (5) and (4).

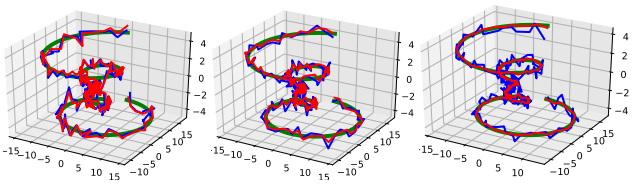
4. EXPERIMENTAL RESULTS

In what follows, we evaluate RSRM on synthetic and real data. The experiments with real fMRI data include adult and developing populations. We use the *Forrest* dataset that includes scans of 18 subjects listening to a 2-hour auditory version of the "Forrest Gump" movie [16]. Each subject contains data from the planum temporale region [17] comprised of 2600 voxels and 3599 TRs. The *audiobook* dataset includes 40 subjects, who listened to a 12 minute (449 TRs) narration of the "Pretty Mouth and Green My Eyes" story from J.D. Salinger [18]. The dataset is divided in two equal-sized groups of 20 subjects. Each group was given a prior context favoring one of two possible interpretations to the story. We use the default mode network ROI [19] comprised of 2500 voxels.

As a new application of RSRM, we use highly variable data acquired from developmental subjects. We collected fMRI data while 6 infant or toddler participants watched a shortened version of a "Mickey's Birthday Party" film for 142 seconds with 4s of burn out (73 TRs). This stimulus was repeated twice for each participant. The age of the subjects ranged from 8 to 36 months: 8, 17, 21, 23, 23, 26, 36. One of the subjects was scanned at 17 and 23 months of age. We define an occipital lobe ROI with an average of 6800 voxels per infant. We used Python to implement the code, and it is released in BrainIAK (`http://brainiak.org`).

4.1. Synthetic data

We begin by evaluating the ability of the algorithm to recover synthetic signals drawn from the model in (1). We create the shared response $\mathbf{R}$ as a three-dimensional curve with 200 TRs long. Then, we generate synthetic signals for 5 subjects by drawing a random projection, $\mathbf{W}^{(i)} \in \mathbb{R}^{30\times 3}$, and projecting the shared response to a 30-dimensional voxel space. Next, the entries of the individual terms $\mathbf{S}^{(i)}$ take a uniformly distributed value in the range $[-4, 4]$ with probability $\rho = 0.2$, and a value 0 with probability $1 - \rho$. Eventually, we add the individual terms and the white Gaussian noise $\mathcal{N}\left(0, \sigma_n^2\right)$ to each subject. The noise level σ_n^2 is chosen following a desired signal-to-noise ratio (SNR). We compare the reconstruction of shared response (curve) with RSRM and SRM for various values of SNR. We set $k = 3$ shared features for both algorithms, and set λ to obtain approximately the same number of non-zeros in the recovered individual terms. We note that the resulting shared response is recovered up to an orthogonal rotation $\mathbf{Q}$ of the space [1]. For the sake of comparison and visualization, we compute and apply $\mathbf{Q}$. Figure 1 depicts the improvement that RSRM achieves over SRM for this model.



(a) 5dB (b) 10dB (c) 20dB

Fig. 1: Examples of Ground truth (green) and reconstructed shared response for synthetic data for RSRM (red) and SRM (blue), when SNR is (a) 5dB, (b) 10dB, and (c) 20dB (λ=1.4, 0.9, and 0.35).

4.2. Time segment matching and group matching on adult participants

Next, we repeat experiments from previous work [4, 1]. In these experiments, we test whether RSRM can identify individual information in the data and in turn better estimate the shared response, hence outperforming SRM. First, we focus on the time segment matching experiment utilizing the *Forrest* dataset. In this experiment, we test whether it is possible to locate a window of 18 seconds (9 TRs) from a held-out subject's data by training a model on the data of the other subjects. The data of each subject is divided into two halves. One half is used for training the functional alignment methods and the other is used to assess performance. Once the model is approximated, we compute the components for the subject that was left out. Next, we transform the other half of the left-out subject's data and look for the highest correlated segment in the averaged shared responses of the other half of the other subjects' data. A match is achieved when the segment is correctly located in the average shared response. Nearby segments are excluded from this computation. We use two-fold cross validation with the halves of the data and leave one out with the subjects.

The results from RSRM are presented in Figure 2 and show a small improvement (about 1.2% for the best performance) over those of SRM. For RSRM we tested several values of λ, and found consistently good results for $\lambda = 0.05$ across the different k values. The small value of λ and the very small improvement in predictive performance, suggests that there

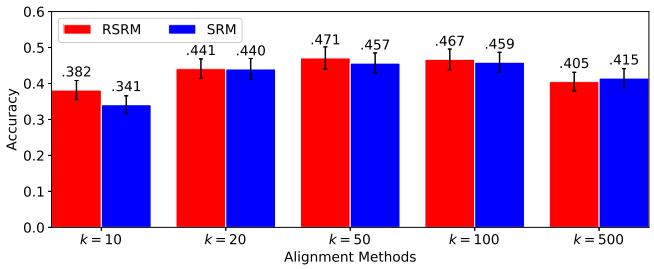


Fig. 2: Time segment matching results with *Forrest* dataset. Accuracy performance is presented for several values of features k for RSRM and SRM. Error bars: ± 1 standard error.

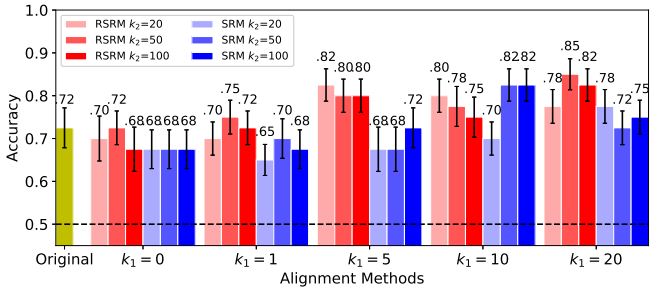


Fig. 3: Group matching results with *audiobook* dataset. Accuracy performance is presented for several values of k_1 inter-group features and k_2 within-group features for RSRM and SRM. Error bars: ± 1 standard error.

is minimal individual information in the data. This may be because the preprocessing of the data in [1] included a voxel selection procedure that selects only task revelant voxels. Even so, RSRM can outperform SRM on these data.

RSRM was also tested with the group matching experiment of [1]. We use the *audiobook* dataset to classify the context of the story to which these participants were exposed. This experiment allows us to evaluate whether data in the individual term can be used to differentiate between groups. We train an SVM classifier on the voxel space data of a set of training subjects while holding-out a subject from each group for testing the classifier. To test functional alignment methods, we first train the algorithms using k_1 features with all the subjects, consequently, we obtain a shared response and a mapping for each subject. Next, we remove the shared response from all subjects, i.e., $\mathbf{Y}^{(i)} = \mathbf{X}^{(i)} - \mathbf{W}^{(i)}\mathbf{R}$. Then, we train two instances of the method with k_2 features on $\mathbf{Y}^{(i)}$'s, using within-group subjects in the training set for each group. We then train the SVM classifier with the shared response projected into the voxel space of each subject, i.e., $\mathbf{W}_g^{(i)}\mathbf{R}_g$, where group $g \in \{1, 2\}$. Testing is performed on the held-out subjects' data.

Figure 3 shows the classification accuracy for both SRM and RSRM. Across most values of k_1 and k_2 RSRM outperforms SRM. The best value for λ was selected within the range $[0.1, 2.0]$. The optimal values of k_1 and k_2 are different between SRM and RSRM. The reason for this is that higher k_1 values allow RSRM to learn more information that is shared

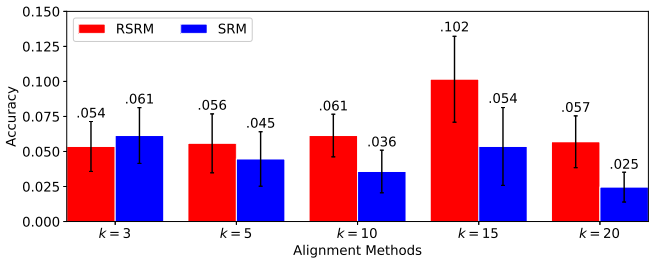


Fig. 4: Time segment matching results with *developmental* dataset. Accuracy performance is presented for several values of features k for RSRM and SRM. Error bars: ± 1 standard error.

while lowering the chance of being influenced by individual data. The results show that the individual information $\mathbf{S}^{(i)}$ can be used to create a shared representation within groups that can discriminate to which contexts participants were exposed.

4.3. Time segment matching on developmental data

In this experiment, we perform time segment matching on the *developmental* data. Following the design of the data collection, we use one repetition of the movie (73 TRs) for training and the other for testing (73 TRs). We locate a window of 9 TRs from a held-out subject on the average shared response of the training group. We apply leave-one-out cross validation on subjects and two-fold cross-validation on each movie repetition.

The results of this experiment are presented in Figure 4 for SRM and RSRM with different values of k. For all values of k, the best value for $\lambda \in [0.01, 3.0]$ is selected for RSRM. The results of the experiment show that RSRM outperforms SRM for values of $k > 3$, improving about 66% over SRM's best accuracy with $k = 15$. When the number of factors k increases, SRM does not improve performance, whereas RSRM is able to maintain or improve the accuracy. RSRM is able to translate part of the variability into the individual terms, allowing for a better shared representation across subjects.

5. CONCLUSIONS

Functional alignment methods have shown their usefulness in aggregating and analyzing fMRI data. However, the performance of these methods can be affected by individual variability. In this work, we presented the robust shared response model that captures a sparse individual information term while simultaneously improving the shared response across subjects. The separation of individual components helped obtain improved mappings as shown in the time segment matching experiment with adult data. The group matching experiment showed that the individual term, which was different across groups, can predict which context participants were exposed to. In infant fMRI data, where the variability of brain functionality across age and subjects is higher than adults, we showed that RSRM captures shared representations better than SRM, boosting predictive power.

6. REFERENCES

[1] P.-H. Chen, J. Chen, Y. Yeshurun, U. Hasson, J. Haxby, and P. Ramadge. A reduced-dimension fmri shared response model. In *NIPS 28*, pages 460–468. 2015.

[2] B. Conroy, B. Singer, J. Haxby, and P. J. Ramadge. fmri-based inter-subject cortical alignment using functional connectivity. In *NIPS 22*, pages 378–386. 2009.

[3] A. G. Huth, W. A. de Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600):453–458, Apr. 2016.

[4] J. Haxby, J. Swaroop Guntupalli, A. Connolly, Y. Halchenko, B. Conroy, M. Gobbini, M. Hanke, and P. Ramadge. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 72(2):404–416, 2011.

[5] G. Varoquaux, A. Gramfort, F. Pedregosa, V. Michel, and B. Thirion. Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In *Information Processing in Medical Imaging*, volume 6801, pages 562–573, Kaufbeuren, Germany, Jul. 2011. Gábor Székely, Horst Hahn, Springer.

[6] S. Khullar, A. Michael, N. Cahill, G. Kiehl, K.and Pearlson, S. Baum, and V. Calhoun. Ica-fnorm: Spatial normalization of fmri data using intrinsic group-ica networks. *Frontiers in Systems Neuroscience*, 5:93, 2011.

[7] Z. Zhou, X. Li, J. Wright, E. Candès, and Y. Ma. Stable principal component pursuit. In *2010 IEEE International Symposium on Information Theory*, pages 1518–1522, Jun. 2010.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, Jun. 2011.

[9] P. Tseng. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.*, 109(3):475–494, Jun. 2001.

[10] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.

[11] K. Vodrahalli, P.-H. Chen, Y. Liang, C. Baldassano, J. Chen, E. Yong, C. Honey, U. Hasson, P. Ramadge, K. Norman, and S. Arora. Mapping Between fMRI Responses to Movies and their Natural Language Annotations. *ArXiv e-prints*, Oct. 2016.

[12] M. J. Anderson, M. Capota, J. S. Turek, X. Zhu, T. L. Willke, Y. Wang, P. H. Chen, J. R. Manning, P. J. Ramadge, and K. A. Norman. Enabling factor analysis on thousand-subject neuroimaging datasets. In *2016 IEEE International Conference on Big Data (Big Data)*, pages 1151–1160, Dec. 2016.

[13] J. S. Turek, T. L. Willke, P.-H. Chen, and P. J. Ramadge. A semi-supervised method for multi-subject fMRI functional alignment. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2017.

[14] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[15] J. C. Gower and G. B. Dijksterhuis. *Procrustes problems*. Oxford University Press, 2004.

[16] M. Hanke, F. J. Baumgartner, P. Ibe, F. R. Kaule, S. Pollmann, O. Speck, W. Zinke, and J. Stadler. A high-resolution 7-tesla fmri dataset from complex natural stimulation with an audio movie. *Scientific Data*, 1:140003 EP –, May 2014.

[17] T. D. Griffiths and J. D. Warren. The planum temporale as a computational hub. *Trends in Neurosciences*, 25(7):348 – 353, 2002.

[18] Y. Yeshurun, S. Swanson, E. Simony, J. Chen, C. Lazaridi, C. J. Honey, and U. Hasson. Same story, different story. *Psychological Science*, 28(3):307–319, 2017. PMID: 28099068.

[19] M. E. Raichle. The brain's default mode network. *Annual Review of Neuroscience*, 38(1):433–447, 2015. PMID: 25938726.