Introduction to MANOVA and CV-MANOVA

Yuanhao Lai*

April 10, 2016

1 MANOVA and CVMANOVA

The method, multivariate analysis of variance (MANOVA) is our method to distinguish different patterns as it is the most well-known traditional multivariate method for comparing different means[1].

1.1 Model Structure

MANOVA assumes our data are generated from the multivariate general linear model (MGLM) [2]„

$$\underset{(n\times p)}{\boldsymbol{Y}} = \underset{(n\times k)(k\times p)}{\boldsymbol{Z}\,\boldsymbol{U}} + \underset{(n\times p)}{\boldsymbol{E}}$$

where $\boldsymbol{Y}$ specifies the signal at n different time points in p voxels, $\boldsymbol{Z}$ is the design matrix with K conditions such as the constant, linear predictors, dummy variables for categorical predictors, $\boldsymbol{U}$ is a matrix of regression coefficients and $\boldsymbol{E}$ is a matrix of random errors. It is assumed that Z is of full column-rank k and each of the n rows of E is an independent sample from a p-dimensional normal distribution $\mathrm{N}(\mathbf{0}, \Sigma_\epsilon)$. Σ_ϵ is the spatial covariance structure of the data. Notice that $n = m \times k$, where m is the number of runs.

The maximum likelihood estimator (MLE) for $\boldsymbol{U}$ is,

$$\hat{\boldsymbol{U}} = \left(\boldsymbol{Z}^T\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}^T\boldsymbol{Y}$$

1.2 Hypothesis Test

We are interested in testing if there is any difference between patterns. The null hypothesis is set as,

$$H_0 : \underset{(q\times k)(k\times p)}{\boldsymbol{C}^T\,\boldsymbol{U}} = \mathbf{0}$$

where $\boldsymbol{C}$ is a $k \times q$ a contrast matrix of the parameters, with rank $q \leq k$.

1.3 MANOVA

To construct a test statistic, we first define the following two important components, which are similar to the uni-variate case.

The regression sum of square,

$$\boldsymbol{Q}_h = \hat{\boldsymbol{U}}^T\boldsymbol{H}^T\boldsymbol{Z}^T\boldsymbol{Z}\boldsymbol{H}\hat{\boldsymbol{U}} \tag{1}$$

where $\boldsymbol{H} = \boldsymbol{C}\left(\boldsymbol{C}^T\boldsymbol{C}\right)^{-1}\boldsymbol{C}^T$ is the hypothesis matrix corresponding to the contrast $\boldsymbol{C}$.

The residual sum of square,

$$\boldsymbol{Q}_e = \left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{U}}\right)^T\left(\boldsymbol{Y} - \boldsymbol{Z}\hat{\boldsymbol{U}}\right) \tag{2}$$

which is also an estimator for the (spatial) covariance structure Σ_ϵ of the data, as $\hat{\Sigma}_\epsilon = \frac{1}{n}\boldsymbol{Q}_e$.

Using $\boldsymbol{Q}_h$ and $\boldsymbol{Q}_e$, we can construct three types of test statistics for MANOVA, namely, the Wilks' lambda, the Pillai's trace and the Barlett-Lawley-Hottellings trace. Those statistics are different in the power in different situations. Normally, Wilks' lambda is used as it is a likelihood ratio test statistic and it is ensured to be the uniformly most powerful test though not the most powerful test in a certain case. Wilks' lambda is defined as below,

*ylai72@uwo.ca, Department of Statistical and Actuarial Sciences, Western University

$$\Lambda = \frac{\det\left(\boldsymbol{Q}_e\right)}{\det\left(\boldsymbol{Q}_h + \boldsymbol{Q}_e\right)} = \prod_i \frac{1}{1+\lambda_i} \tag{3}$$

where λ_i are the solutions of the characteristic equation $\mid \boldsymbol{Q}_h - \lambda \boldsymbol{Q}_e \mid = 0$.

In addition, the F approximation [3] to the null distribution of Wilks's Lambda is available,

$$F = \left[\Lambda^{\frac{-1}{t_3}} - 1\right] \cdot \frac{\mathrm{df}_2}{\mathrm{df}_1} \sim F_{\mathrm{df}_1, \mathrm{df}_2}$$

where $t_1 = n - k - \frac{p-q+1}{2}$, $t_2 = \frac{pq-2}{4}$, $t_3 = \begin{cases} \sqrt{\frac{(pq)^2-4}{p^2+q^2-5}}, & p^2+q^2-5 > 0 \\ 1, & p^2+q^2-5 \leq 0 \end{cases}$, $\mathrm{df}_1 = pq$ and $\mathrm{df}_2 = t_1 \cdot t_3 - 2t_2$.

Notice that the approximated F distribution fails when $n-k \geq p$ because df_2 may be less than 0. The essential reason is that $\boldsymbol{Q}_e$ defined by eqn.(2) is not positive-definite in this case and hence the computation of the statistic is not feasible. To overcome this problem, we proposed using the principle component analysis to reduce the number of voxels. We would discuss it later in the Result section.

1.4 CV-MANOVA

In the case where we want to quantify the distinctness between patterns, we prefer the Barlett-Lawley-Hottellings trace because its equivalent relationship to the Mahalanobis distance as shown by Allefeld [4]. However, the Barlett-Lawley-Hottellings trace is severely biased, so Allefeld used the cross-validated version of the Barlett-Lawley-Hottellings trace to make it unbiased. The idea is very simple. He replaced the estimated pattern $\hat{U}$ in eqn.(1) with its cross-validated version. This alternative approach under the same MGLM framework is called the cross-validated MANOVA (CV-MANOVA). The resulted statistic by the 'leave-one-run-out' cross-validation is,

$$D = \frac{(m-1)(n-k)-p-1}{(m-1)n} \cdot \frac{1}{m} \sum_{l=1}^{m} \sum_{h \neq l} \mathrm{trace}\left[\boldsymbol{Q}_h^{CV} \left(\boldsymbol{Q}_e\right)^{-1}\right] \tag{4}$$

where $\boldsymbol{Q}_h^{CV} = \hat{\boldsymbol{U}}^{(h)T} \boldsymbol{H}^T \boldsymbol{Z}^{(l)T} \boldsymbol{Z}^{(l)} \boldsymbol{H} \hat{\boldsymbol{U}}^{(l)}$ and $\hat{\boldsymbol{U}}^{(h)}$ is the MLE of $\boldsymbol{U}$ using only the hth run data. However, CV-MANOVA suffered the same problem as MANOVA for $\boldsymbol{Q}_e$ when $n-k \geq p$. But since our simulated data is assumed to be (spatially) prewhitened , $\boldsymbol{Q}_e = \hat{\Sigma}_\epsilon$ can be estimated as a diagonal matrix where the diagonal elements are the same as eqn.(2). However, there is not an exact distribution or a good approximated distribution for this statistic. So the distribution under H_0 of the CV-MANOVA is obtained by either simulations or permutations.

2 Relationship between CV-MANOVA and Pattern Component Modeling

It can be shown the CV-MANOVA D is also a linear contrasts of the second moment matrix $\hat{G}$ [5].. If we ignored a constant scale of the cross-validated Barlett-Lawley-Hottellings trace in eqn.(4), then

$$\begin{aligned}
D &= \sum_{l=1}^{m} \sum_{h \neq l} \mathrm{trace}\left[\boldsymbol{Q}_h^{CV} \left(\boldsymbol{Q}_e\right)^{-1}\right] \\
&= \sum_{l=1}^{m} \sum_{h \neq l} \mathrm{trace}\left[\hat{\boldsymbol{U}}^{(h)T} \boldsymbol{H}^T \boldsymbol{Z}^{(l)T} \boldsymbol{Z}^{(l)} \boldsymbol{H} \hat{\boldsymbol{U}}^{(l)} \left(\hat{\Sigma}_\epsilon\right)^{-1}\right] \\
&= \sum_{l=1}^{m} \sum_{h \neq l} \mathrm{trace}\left[\hat{\boldsymbol{U}}^{(h)T} \boldsymbol{H}^T \boldsymbol{Z}^{(l)T} \boldsymbol{Z}^{(l)} \boldsymbol{H} \hat{\boldsymbol{U}}^{(l)} \left(\hat{\Sigma}_\epsilon\right)^{-1} \hat{\boldsymbol{U}}^{(h)T}\right], \text{ by the property of matrix trace} \\
&= (m-1)m \times \mathrm{trace}\left[\boldsymbol{H}^T \boldsymbol{Z}^{(l)T} \boldsymbol{Z}^{(l)} \boldsymbol{H} \hat{G}\right] \\
&\Rightarrow \mathrm{trace}\left[\boldsymbol{H}^T \boldsymbol{H} \hat{G}\right], \text{ if the design is balanced,} \boldsymbol{Z}^{(l)T} \boldsymbol{Z}^{(l)} \approx cI \\
&= \mathrm{trace}\left[\boldsymbol{H} \hat{G}\right], \ \boldsymbol{H} \text{ is idempotent and symmetric} \\
&= \sum_{i,j} \boldsymbol{H}_{ij} \hat{G}_{ij}
\end{aligned}$$

Both CV-MANOVA and PCM estimate $\hat{\boldsymbol{U}}$ for all terms simultaneously, so they have the same $\hat{G}$. Therefore, the equivalence between CV-MANOVA and PCM require using the same hypothesis matrix $\boldsymbol{H}$ and $Z^T Z \approx cI$, where c is a constant. However, $\boldsymbol{H}$ is not necessarily the same for CV-MANOVA and PCM. Although $\boldsymbol{H}$ for both methods are based on the same contrast matrix, the construction procedures from the contrast matrix to $\boldsymbol{H}$ are different. For example, if we are interested in testing if any of the main effects, the interaction and the overall (intercept) effect are significant, then we will have four contrast matrices $C_{\text{main 1}}$, $C_{\text{main 2}}$, $C_{\text{interaction}}$ and $C_{\text{intercept}}$ and we will need four hypothesis matrices based on them. But CV-MANOVA forms four $\boldsymbol{H}$'s independently each contrast matrix, while PCM forms four $\boldsymbol{H}$'s from the combinations of all contrast matrices. If we are only interested in testing $C_{\text{main 1}}$ and $C_{\text{main 2}}$, then $\boldsymbol{H}$'s formed by the CV-MANOVA and PCM will be the same because the $C_{\text{main 1}}$ and $C_{\text{main 2}}$ are linearly independent, but this is not true when $C_{\text{interaction}}$ is involved. From this point of view, we may prefer PCM though they should have similar performances.

3 Results

For all the statistical tests we performed, we specified type I error i as 0.05. So the critical value for each method was obtained as the 95th percentile of the simulated/permuted distribution (one-sided test).

3.1 Limitation of MANOVA

As we mentioned in Section 1.3, MANOVA fails when $n - k \geq p$. To overcome this problem, we introduced two simple methods and used simulations to evaluate them . For the first method, we only utilized the first 45 voxels and drop the remaining voxels. Figure 1 shows that the F approximation could still control the false positive rate while retaining some powers. But this method was deficient as it discarded a large amount of the original signals. A better approach was to use the principal component analysis to reduce the dimensions of voxels. Figure 2 shows that for the number of components between 25 to 30, the F approximation also controls the false positive rate and achieves a much higher power. Figure 3 also shows that the explained variances by these components are around 70%.

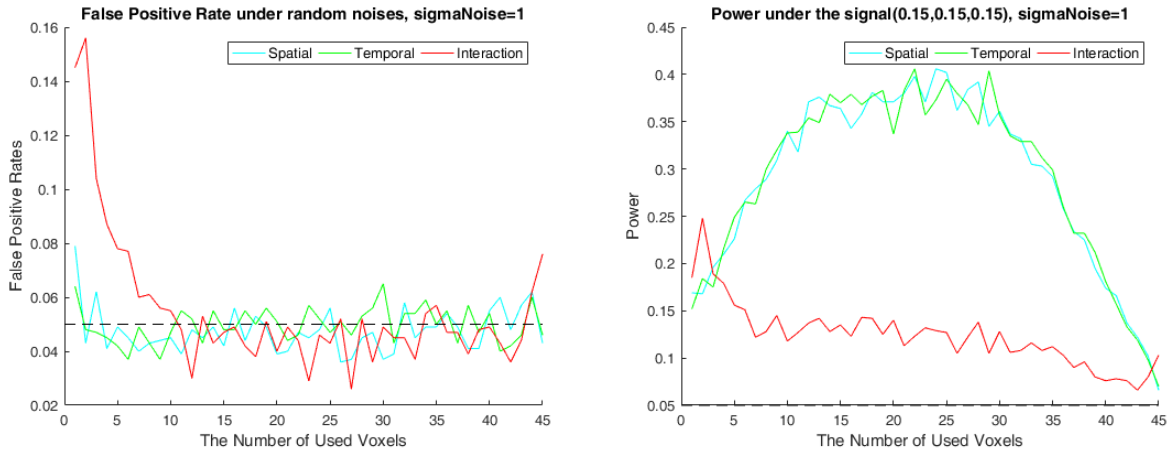

Figure 1: MANOVA Using the first N voxels ($N = 1, \ldots, 45$): The left figure shows the false positive rates for detecting each term from 1000 simulations of random noises. The right figure shows the power for detecting each term from 1000 simulations of signals, with 0.15 scale for each term. Under both cases, the variance for the random error term is 1.

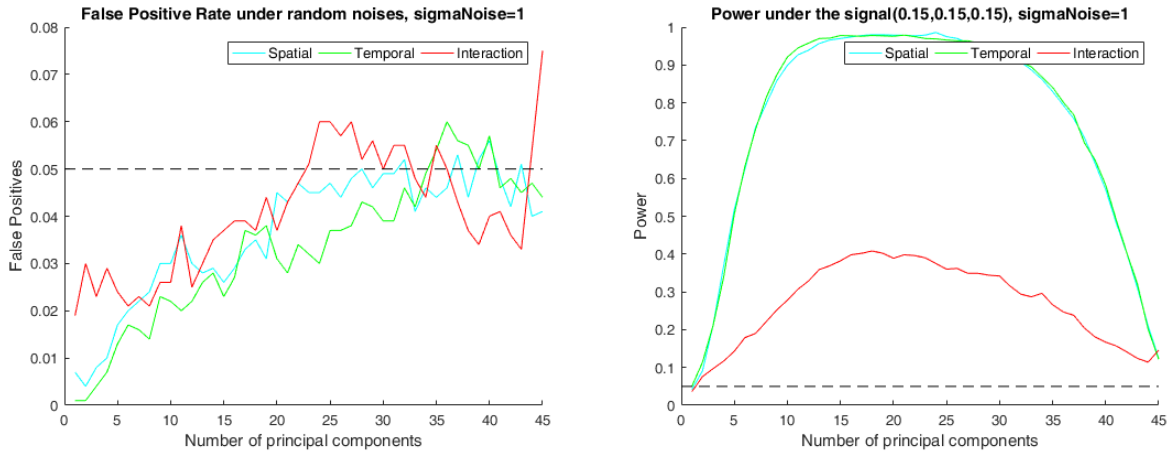

Figure 2: CV-MANOVA using the first N principal components of voxels: The left figure shows the false positive rates for detecting each term from 1000 simulations of random noises. The right figure shows the power for detecting each term from 1000 simulations of signals, with 0.15 scale for each term. Under both cases, the variance for the random error term is 1. ($N = 1, \ldots, 45$)

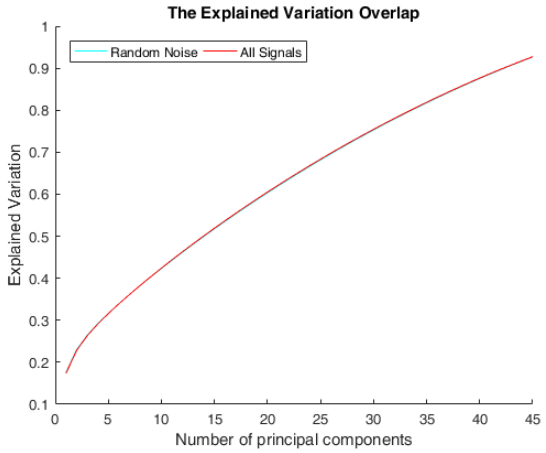

Figure 3: Explained Variance of PCA

3.2 Power under Different Hypotheses

To compare the power of different methods under different hypothese, we did 1000 simulations under different hypothesis and the results are shown by Figure 4 to Figure 5. The three digits in the legend represents if there is an effect of the spatial pattern, the temporal pattern or the interaction, where 0 indicates no such patterns and 1 indicates the signal for the pattern is 0.15. The curves with no markers were set to the null distribution for performing statistical tests, because the the pattern of interest did not exist. Hence the numbers besides the curves with markers indicate the power for detecting the pattern in the corresponding alternative. We can see that the CV-MANOVA enhanced the powers for both main effects and interactions. We also found out that the CV-MANOVA was approximately unbiased when the data is generated from random noises, but there was a drift for the distribution of the spatial pattern when the interaction was involved.

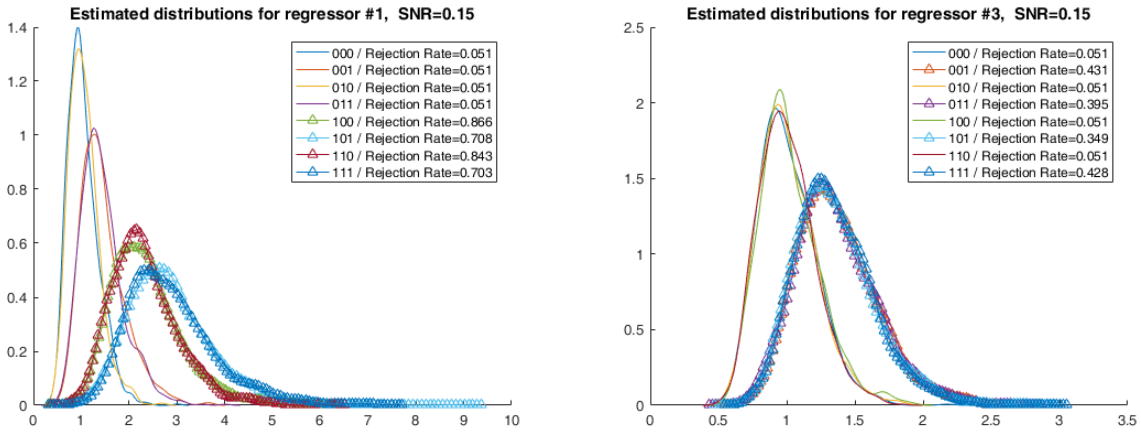

Figure 4: MANOVA using the first 28 principal components under difference hypotheses: the left figure shows the power for detecting the spatial pattern and the right figure is for the interaction.

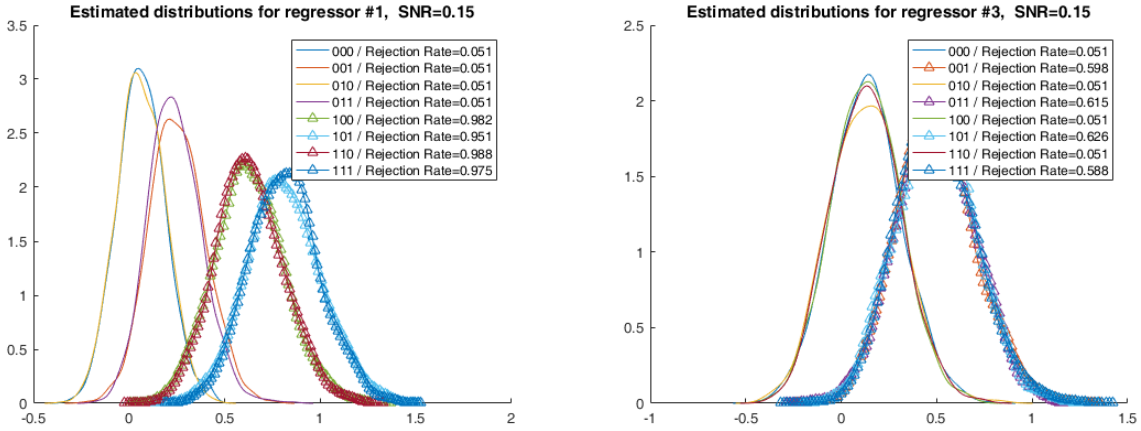

Figure 5: CV-MANOVA under difference hypotheses: the left figure shows the power for detecting the spatial pattern and the right figure is for the interaction.

3.3 Noise Tolerance of MANOVA and CVMANOVA

In order to see how the power of each method is related the signal-noise-ratio (SNR), we did 1000 simulations with four different SNR's. The null hypothesis for each method was set as random noises. Figure 6 to Figure 7 show that the power of each method increased dramatically as the SNR increased.

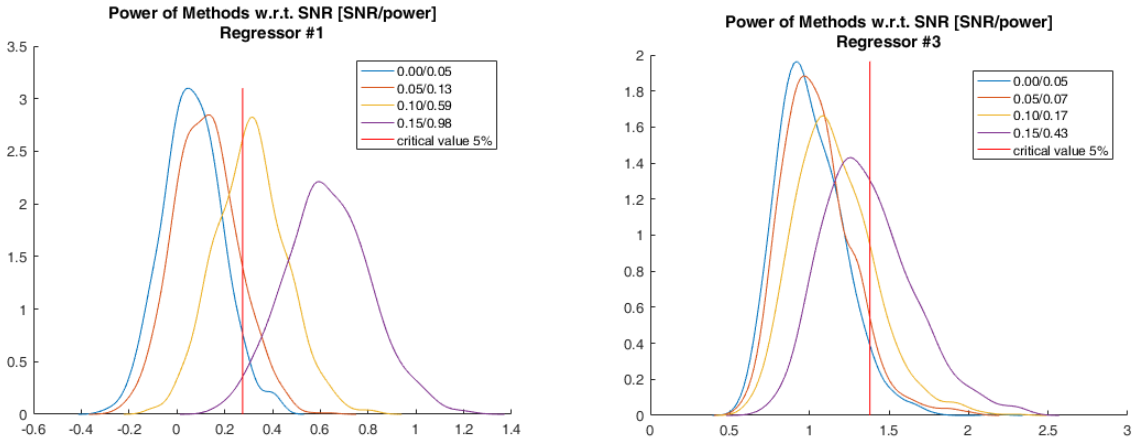

Figure 6: Power v.s. SNR for MANOVA using the first 28 principal components under difference hypotheses.

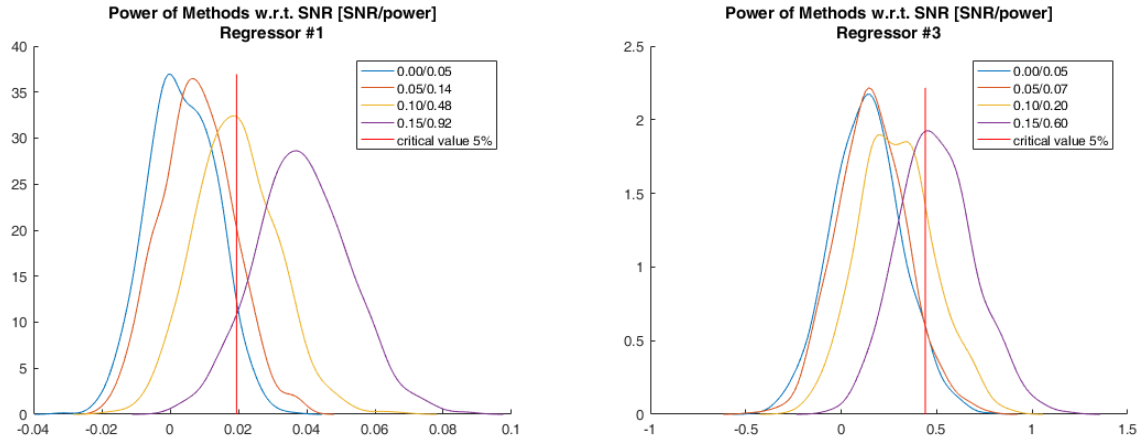

Figure 7: Power v.s. SNR for CVMANOVA

References

[1] Charles S Davis. *Statistical methods for the analysis of repeated measurements*. Springer Science & Business Media, 2002.

[2] John Fox, Michael Friendly, and Sanford Weisberg. Hypothesis tests for multivariate linear models using the car package. *R J*, 5(1):39–52, 2013.

[3] C Radhakrishna Rao. An asymptotic expansion of the distribution of wilks¡¯ criterion. *Bulletin of the International Statistical Institute*, 33(Part 2):177–180, 1951.

[4] Carsten Allefeld and John-Dylan Haynes. Searchlight-based multi-voxel pattern analysis of fmri by cross-validated manova. *Neuroimage*, 89:345–357, 2014.

[5] Alexander Walther, Naveed Ejaz, Nikolaus Kriegeskorte, Jörn Diedrichsen, and Jörn Diedrichsen. Representational fmri analysis: an introductory tutorial.

[6] James V Haxby. Multivariate pattern analysis of fmri: the early beginnings. *Neuroimage*, 62(2):852–855, 2012.

[7] Alexander Walther, Hamed Nili, Naveed Ejaz, Arjen Alink, Nikolaus Kriegeskorte, and Jörn Diedrichsen. Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, 2015.

[8] Katja Kornysheva and Jörn Diedrichsen. Human premotor areas parse sequences into their spatial and temporal features. *Elife*, 3:e03043, 2014.

[9] Seiji Ogawa, Tso-Ming Lee, Alan R Kay, and David W Tank. Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences*, 87(24):9868–9872, 1990.

[10] Amir Shmuel, Essa Yacoub, Josef Pfeuffer, Pierre-Francois Van de Moortele, Gregor Adriany, Xiaoping Hu, and Kamil Ugurbil. Sustained negative bold, blood flow and oxygen consumption response and its coupling to the positive response in the human brain. *Neuron*, 36(6):1195–1210, 2002.

[11] Scott A Huettel, Allen W Song, and Gregory McCarthy. *Functional magnetic resonance imaging*, volume 1. Sinauer Associates Sunderland, 2004.

[12] James V Haxby, M Ida Gobbini, Maura L Furey, Alumit Ishai, Jennifer L Schouten, and Pietro Pietrini. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293(5539):2425–2430, 2001.

[13] Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:4, 2008.