Prewhitening High Dimensional fMRI Data Sets Without Eigendecomposition

Abd-Krim Seghouane[1] and Yousef Saad[2]

Department of Electrical and Electronic Engineering, Melbourne School of Engineering

The University of Melbourne, Melbourne, Australia

E-mail: Abd-krim.seghouane@unimelb.edu.au

[2] Department of Computer Science and Engineering,

The University of Minnesota at Twin Cities, Minnesota, USA

Email: saad@umn.edu

Abstract

This letter proposes an algorithm for linear whitening that minimizes the mean squared error (MSE) between the original and whitened data without using the truncated eigen decomposition (ED) of the covariance matrix of the original data. This algorithm uses Lanczos vectors to accurately approximate the major eigenvectors and eigenvalues of the covariance matrix of the original data. The major advantage of the proposed whitening approach is its low computational cost when compared with that of the truncated ED. This gain comes without sacrificing accuracy as illustrated on an experiment of whitening a real high dimensional fMRI data set.

Keywords: fMRI, temporal ICA, Krylov subspace, Lanczos vectors, eigendecomposition.

1 Introduction

The univariate and multivariate linear regression models are simple and widely used parametric models for fMRI data analysis Ashby (2011), Lazar (2010). In these models the linear (deterministic) part is used to characterize the activation response in a single voxel for the univariate model or a group of voxels for the multivariate model and the baseline drift whereas the second (stochastic) part of the model characterizes the noise from both physical and physiological processes Ashby (2011), Lazar (2010).

On the one hand, the univariate linear regression model uses a hemodynamic response function (HRF) as a parameter vector with extra parameters for the drift and a univariate temporally correlated noise to model the fMRI time series at each voxel to infer task-related activations with estimates of the level of significance Ardekani & Kershaw & Kashikura & Kanno (1999), Seghouane & Shah (2012). As a consequence, the sensitivity of hypothesis methods based on this model across a group of voxel is crucially dependent on the group correction post processing step Genovese & Lazar & Nichols (2002).

On the other hand, multivariate linear regression models have mainly been used in data-driven methods Esposito & al. (2002), Calhoun & Adali & Pearlson & Pekar (2001). They allow the exploitation of the relationships between voxels. Among the multivariate data-driven techniques used in fMRI, independent component analysis (ICA) Hyvarien & Karhunen & Oja (2001), Jolliffe (2002) has been widely applied to fMRI to find components that are spatially independent Esposito & al. (2002) (sICA) or temporally independent Calhoun & Adali & Pearlson & Pekar (2001) (tICA).

A common pre-processing step used with both the above models is prewhitening. In the case of the univariate model, this allows the generation of the best linear unbiased estimates of the parameters vector and therefore more accurate activation tests. In the ICA case, it has the advantage to reduce the complexity and improve the convergence of the ICA algorithm Hyvarien

& Karhunen & Oja (2001).

Prewhitening is obtained by multiplying the fMRI data by the square root of the inverse covariance matrix which can be obtained from the Cholesky factorization or the eigen decomposition of the inverse covariance matrix. As it is well known, however, the Cholesky factorization or the eigen decomposition are computationally expensive, scaling $O(p^3)$ for a dense $p \times p$ covariance matrix. While this is not a major inconvenience when dealing with univariate linear regression modeling of fMRI time series or sICA, this can be a major problem for tICA when the number of voxels considered is very large. In fMRI analysis, the spatial resolution is often at least about 64×64 over 32 slices resulting in 131072 voxels. Prewhitening needs substantial computational work to get the eigen decomposition, to overcome this inconvenient an alternative approach for prewhitening large dimensional multivariate vectors is proposed in this paper. The proposed subspace method constructs an approximated whitened vector based on Krylov sequences of subspaces reachable from the unwhitened vector. The goal of the proposed algorithm is identical with the approach based on ED, namely to preserve the quality of the resulting product between the inverse square root of the covariance matrix and the data vector to be whitened in the major eigen directions of the covariance matrix. A Lanczos based approach is used to achieve this goal by using a relatively small number r of Lanczos vectors. The advantage of the proposed approach is its computational complexity which is $O(rp^2)$ with $r \ll p$ compared with the $O(p^3)$ associated with the ED-based approach. Like the Cholesky factorization or the eigen decomposition, the proposed method involves an eigen decomposition (factorization) but it is implemented on a much smaller matrix resulting in a much computationally cheaper whitening method in comparison to $O(p^3)$. The proposed method is particularly appealing when a reduced rank approximation of the covariance matrix is used for prewhitening.

The rest of the paper is organized as follows: the prewhitening preprocessing step of multivariate temporal fMRI time series is reviewed in the next section. The proposed prewhitening algorithm is described in Section 3. The choice of the dimension of the krylov subspace is discussed in

Section 4. The performance of the proposed whitening method on real fMRI data is illustrated in Section 5. Concluding remarks are given in Section 6.

2 Prewhitening and multivariate fMRI temporal model

ICA has become the main model for multivariate data-driven fMRI analysis. With this model, an observed data vector $\mathbf{y} = (y_1, y_2, ..., y_p)^\top$ is modeled as a mixture of an unobservable source vector $\mathbf{x} = (x_1, x_2, ..., x_m)^\top$ as follows

$$\mathbf{y} = A\mathbf{x} \tag{1}$$

where $A \in R^{p\times m}$ is a mixing matrix. In tICA, a measurement matrix $Y = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n) \in R^{p\times n}$ is formed by collecting fMRI times series of time length n across p voxels. The matrix $X \in R^{m\times n}$ contains m temporally independent sources of length n. In the prewhitening step, the mixed vector $\mathbf{y}$ is processed by

$$\mathbf{z} = V^{-\frac{1}{2}}(\mathbf{y} - \mu) \tag{2}$$

where μ and V are the mean vector and covariance matrix of $\mathbf{y}$, such that $\mathbf{z}$ is $N(0, \sigma^2 I_p)$. In the classical prewhitening, they are estimated by

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i \quad \text{and} \quad \hat{V} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{y} - \hat{\mu})(\mathbf{y} - \hat{\mu})^\top$$

in a batch way sampling.

A prewhitened data set is necessary in some ICA algorithms. It reduces the complexity of the ICA problems Hyvarien & Karhunen & Oja (2001). Given V assumed full rank positive definite, $V^{-\frac{1}{2}}$ is usually obtained from the eigen decomposition as follows

$$V^{-\frac{1}{2}} = US^{-\frac{1}{2}}U^\top \tag{3}$$

where $S = \text{diag}(s_1, s_2, ..., s_p)$ is the diagonal matrix of eigenvalues and U the unitary matrix of eigenvectors. However, as in many scientific fields including tICA in fMRI, where the number of samples n is much smaller than the number of variables p, although the data are presented in a high dimensional space, they actually have a much lower intrinsic dimensionality $k \ll p$. In this case V is generally rank deficient and the best full rank-k approximation of V in the 2-norm (obtained by truncating the ED) is used instead of V

$$V_k^{-\frac{1}{2}} = U_k S_k^{-\frac{1}{2}} U_k^\top \tag{4}$$

where U_k consists of the first k columns of U and S_k is the k-th principal submatrix of S. When a very large number of voxels is considered, prewhitening is computationally expensive since it requires a computational complexity of order $O(np^2)$ for the covariance matrix and $O(p^3)$ for the eigen decomposition. The usefulness of the later is questionable since only the major eigenvalues and eigenvectors of V are needed in (4). Furthermore, the least-squares residual associated to the centered whitened vector with (4) is given by

$$\begin{aligned} R_k &= ||V^{\frac{1}{2}}\mathbf{z} - \mathbf{y}||^2 \\ &= ||V^{\frac{1}{2}} U_k S_k^{-\frac{1}{2}} U_k^\top \mathbf{y} - \mathbf{y}||^2 \\ &= ||U_k^\top \mathbf{y} - U^\top \mathbf{y}||^2 \\ &= \sum_{i=k+1}^{p} (\mathbf{u}_i^\top \mathbf{y})^2 \end{aligned} \tag{5}$$

Relation (5) shows that the truncated ED used for choosing the order in which to include components is unnatural. Based on the least squares residual, the components should be chosen using the magnitude of the components of $U^\top \mathbf{y}$. The singular values (principal components) are ordered in such a way that they reflect the directions of largest variation in the covariance matrix V. However, the purpose of the whitening procedure is to reduce the least squares residual using a smaller dimension projection subspace, then the choice of the components should be done according to the components of the vector $U^\top \mathbf{y}$.

In what follows, based on Krylov sequences of subspaces reachable from the vector $\mathbf{y}$, a computationally efficient approach is proposed for generating the major eigenvectors and eigenvalues of V without computing its eigendecomposition. An eigendecomposition is still necessary but is applied on a much smaller matrix of dimension $r \ll p$.

3 Krylov subspace approximations for prewhitening

Given a symmetric matrix $V \in R^{p\times p}$ and an initial vector **y**, a Krylov subspace is a subspace of the form

$$K_r(V, \mathbf{y}) = \text{span}\left\{\mathbf{y}, V\mathbf{y}, V^2\mathbf{y}, ..., V^{r-1}\mathbf{y}\right\}.$$

Increasing the dimension r allow us to reach the entire subspace of the pair $(V, \mathbf{y})$. The dimension $r \leq p$. In our computation we not only use the Krylov subspace $K_r(V, \mathbf{y})$ but also an orthonormal basis of it. The Lanczos algorithm builds such orthonormal basis.

The Lanczos algorithm generates a set of r vectors q_i, $i = 1, ..., r$ that form an orthonormal basis of the Krylov subspace $K_r(V, \mathbf{y})$. These vectors satisfy the 3-term recurrence

$$\beta_{i+1}q_{i+1} = Vq_i - \alpha_i q_i - \beta_i q_{i-1}$$

with $\beta_1 q_0 = 0$. The coefficients α_i and β_{i+1} are computed so as to ensure that $\langle q_{i+1}, q_i\rangle = 0$ and $||q_{i+1}|| = 1$. The time cost of this algorithm is $O(rp^2)$ for a dense matrix V. If $Q_r = [q_1, q_2, ..., q_r] \in R^{p\times r}$, then an important equality that results from the algorithm is

$$Q_r^\top V Q_r = T_r = \begin{bmatrix} \alpha_1 & \beta_2 & & & \\ \beta_2 & \alpha_2 & \beta_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \beta_{r-1} & \alpha_{r-1} & \beta_r \\ & & & \beta_r & \alpha_r \end{bmatrix}. \tag{6}$$

where T_k is a tridiagonal matrix. The eigenvalues of T_r are called Ritz values and for an eigenvector **w** of T_r, $Q_r\mathbf{w}$ defines the Ritz vector. As r increases more Ritz values and vectors will converges towards the eigenvalues and eigenvectors of V Golub & Van Loan (1996), Saad (1992), Chen & Saad (2009).

The Krylov subspace $K_r(V, \mathbf{y})$ reached from a small number of r of naive vectors is used for the construction of an approximation of **z**. The eigen decomposition of T_r defined in (6) obtained from $K_r(V, \mathbf{y})$ is easily computed in practice because r is assumed to be small

$$T_r = WDW^\top \quad \text{and} \quad T_r^{-\frac{1}{2}} = WD^{-\frac{1}{2}}W^\top. \tag{7}$$

The eigenvalues in D are known to approximate the extremal eigenvalues of V and $Q_r W$ are the approximate eigenvectors.

The problem of linear whitening can be formulated as searching for the vector **z** such that Eldar & Oppenheim (2005)

$$E\left\{(\mathbf{z} - \mathbf{y})^2\right\} \tag{8}$$

is minimum subject to

$$E\left\{\mathbf{z}\mathbf{z}^\top\right\} = \sigma^2 I_p. \tag{9}$$

Assuming the vector $\mathbf{z} \in K_r(V, \mathbf{y})$, hence it can be written as

$$\mathbf{z} = K_r \mathbf{a} \tag{10}$$

for some appropriate $\mathbf{a} \in R^r$. The orthonormal basis Q_r generated by the Lanczos algorithm is used to represent this vector. In the basis Q_r, $\mathbf{z} \in K_r(V, \mathbf{y})$ can be written as

$$\hat{\mathbf{z}} = Q_r \mathbf{b} = Q_r W W^\top \mathbf{b} \equiv U\mathbf{t} \tag{11}$$

for some appropriate $\mathbf{b} \in R^r$ and **t** is $N(0, \sigma^2 I_r)$ since U is orthogonal. With this the MSE (8) above can be written as

$$\begin{aligned} E\left\{(\mathbf{z} - \mathbf{y})^2\right\} &= E\left\{(\mathbf{t} - U^\top \mathbf{y})^2\right\} \\ &= r\sigma^2 + \sum_{i=1}^{r} s_i - 2\sum_{i=1}^{r} E\left(t_i (U^\top y)_i\right) \\ &\simeq r\sigma^2 + \sum_{i=1}^{r} d_i - 2\sum_{i=1}^{r} E\left(t_i (U^\top y)_i\right) \end{aligned} \tag{12}$$

where t_i is $N(0, \sigma^2)$ and $(U^\top y)_i$ is $N(0, s_i)$. the minimum of (12) is obtained for the maximum of the last term of the right hand side of (12). Using Cauchy-Schwartz inequality

$$E\left(t_i(U^\top y)_i\right) \leq \sqrt{\text{var}(t_i)\text{var}((U^\top y)_i)} = \sigma\sqrt{\text{var}((U^\top y)_i)} \tag{13}$$

the maximum of (12) is obtained for

$$t_i = \sigma\frac{(U^\top y)_i}{\sqrt{s_i}} \simeq \sigma\frac{(U^\top y)_i}{\sqrt{d_i}}. \tag{14}$$

Therefore

$$\mathbf{b} \simeq \sigma W D^{-1/2} W^\top Q_r^\top \mathbf{y} = \sigma T_r^{-1/2} Q_r^\top \mathbf{y} \tag{15}$$

and

$$\mathbf{z} \simeq \sigma Q_r T_r^{-1/2} Q_r^\top \mathbf{y}. \tag{16}$$

By replacing $\mathbf{t} = \sigma D^{-1/2} U^\top \mathbf{y}$ in (12) and minimizing with respect to σ, we have

$$\hat{\sigma} = \frac{1}{r}\sum_{i=1}^{r}\sqrt{d_i}. \tag{17}$$

Using (17), **z** becomes

$$\mathbf{z} \simeq \hat{\sigma}\beta Q_r T_r^{-\frac{1}{2}} e_1 \tag{18}$$

where $\beta = ||\mathbf{y}||_2$, e_1 is the first column of the $r \times r$ identity matrix and the first vector of the basis Q_r is $q_1 = \mathbf{y}/||\mathbf{y}||_2$. The approximate whitened vector is then based on computing the inverse square root of a matrix of substantially smaller dimension $r \ll p$, where it is inexpensive to compute exactly.

The above procedure is captured in the algorithm below.

Algorithm: Lanczos procedure for whitening

Input: V, r, **y**

1. Set $\beta_1 \leftarrow 0$, $q_0 \leftarrow 0$.

2. Choose $q_1 = \mathbf{y}/||\mathbf{y}||_2$.

3. **For** i=1,...,r **do**

4. $l_i \leftarrow Vq_i - \beta_i q_{i-1}$

5. $\alpha_i \leftarrow (l_i, q_i)$

6. $l_i \leftarrow l_i - \alpha_i q_i$

7. $\beta_{i+1} \leftarrow ||l_i||_2$

8. $q_{i+1} \leftarrow l_i/\beta_{i+1}$

9. **EndFor**

10. Set $Q_r \leftarrow [q_1, q_2, ..., q_r]$, $T_r \leftarrow$ tridiag$\{\beta_j, \alpha_j, \beta_j\}$.

11. Compute the eigendecomposition of $T_r = WDW^\top$

12. Set $\hat{\sigma} \leftarrow \frac{1}{r}\sum_{i=1}^{r}\sqrt{d_i}$.

13. **Return** $\mathbf{z} = \hat{\sigma}\beta Q_r T_r^{-\frac{1}{2}} e_1$

Since the whitened vector of interest is of the form $\mathbf{z} = V_k^{-\frac{1}{2}}\mathbf{y}$, the vector $\hat{\mathbf{z}}$ defined in (18) is a good alternative. Using

$$
\begin{aligned}
T_r^{\frac{-1}{2}} = WD^{\frac{-1}{2}}W^\top &= Q_r^\top Q_r W D^{-\frac{1}{2}} W^\top Q_r^\top Q_r \\
&\simeq Q_r^\top V^{-\frac{1}{2}} Q_r
\end{aligned} \tag{19}
$$

where the last term is valid for sufficiently large r, in (18) gives

$$
\hat{\mathbf{z}} = \hat{\sigma} Q_r Q_r^\top V^{-\frac{1}{2}}\mathbf{y} \tag{20}
$$

which is the orthogonal projection of the exact solution (2) onto the Krylov subspace. When projected to the subspace spanned by the major eigenvectors of V, this approximation is very close to the projection of $\mathbf{z}$ if r is sufficiently large.

The proposed approach (16) as well as the truncated ED (4) are two methods for biased estimation. Each of these methods generate a filtered version of $\mathbf{z}$ given in (2). In the case of the truncated ED we have

$$
\mathbf{z}_1 = V_k^{-\frac{1}{2}}\mathbf{y} = \sum_{i=1}^{p} f_1(s_i)\frac{\mathbf{u}_i^\top \mathbf{y}}{\sqrt{s_i}}\mathbf{u}_i \tag{21}
$$

where the shrinkage function $f_1(s_i)$ is given by

$$f_1(s_i) = \begin{cases} 1 & \text{for} \quad 1 \leq i \leq k \\ 0 & \text{for} \quad k < i \leq p \end{cases}$$

The bias is introduced in the directions that are responsible for the high variance. The hope is that the increase in bias is small compared to the decrease in variance leading to a reduction in the mean square error. If the shrinkage function $f_1(.)$ above does not depends on **y**, that is $\hat{\mathbf{z}}$ is linear in **y**, any $f_1(s_i) \neq 1$ will increase the bias in the ith direction. The proposed whitening procedure in (16) is a filtering estimator as well. The shrinkage function in this case is closely related to the Ritz values and vectors and is given by Sluis & Vorst (1997)

$$f_2(s_i) = 1 - \prod_{j=1}^{r} \left(1 - \frac{\sqrt{s_i}}{\sqrt{d_j}}\right).$$

In comparison to (21),

$$\mathbf{z}_2 = \sigma Q_r T_r^{-\frac{1}{2}} Q_r^\top \mathbf{y} = \sum_{i=1}^{p} f_2(s_i) \mathbf{u}_i^\top \mathbf{y} \mathbf{u}_i$$

is nonlinear in **y**. The values $f(s_i)$ depend on the eigenvalues of T_r and T_r in turn depends on **y** in a complicated manner. f_2 is related to (5) through $K_1, ..., K_r$ which are subspaces of the space spanned by the vectors $\mathbf{u}_i$ for which $\mathbf{u}_i^\top \mathbf{y} \neq 0$. Q_r is obtained through a Gram-Schmidt orthonormalization (GSO) of K_r

$$\begin{aligned} Q_r = GSO\left\{K_r(V, \mathbf{y})\right\} &= GSO\left\{\mathbf{y}, V\mathbf{y}, V^2\mathbf{y}, ..., V^{r-1}\mathbf{y}\right\} \\ &= GSO\left\{UU^\top\mathbf{y}, USU^\top\mathbf{y}, US^2U^\top\mathbf{y}, ..., US^{r-1}U^\top\mathbf{y}\right\} \\ &= U.GSO\left\{U^\top\mathbf{y}, SU^\top\mathbf{y}, S^2U^\top\mathbf{y}, ..., S^{r-1}U^\top\mathbf{y}\right\}. \end{aligned}$$

The values of $f_2(s_i)$ automatically take into account both (5) and the variances in V whereas the value of f_1 extract only the directions $\mathbf{u}_i$ that explain the maximum variance in V.

The convergence of $\hat{\mathbf{z}}$ to **z** in the directions of the major eigenvectors of V as r increases can be analyzed using

$$|\langle(\mathbf{z} - \hat{\mathbf{z}}), \mathbf{u}_j\rangle| \leq c_j ||\mathbf{z}||_2 T_m^{-1}(\gamma_j) \tag{22}$$

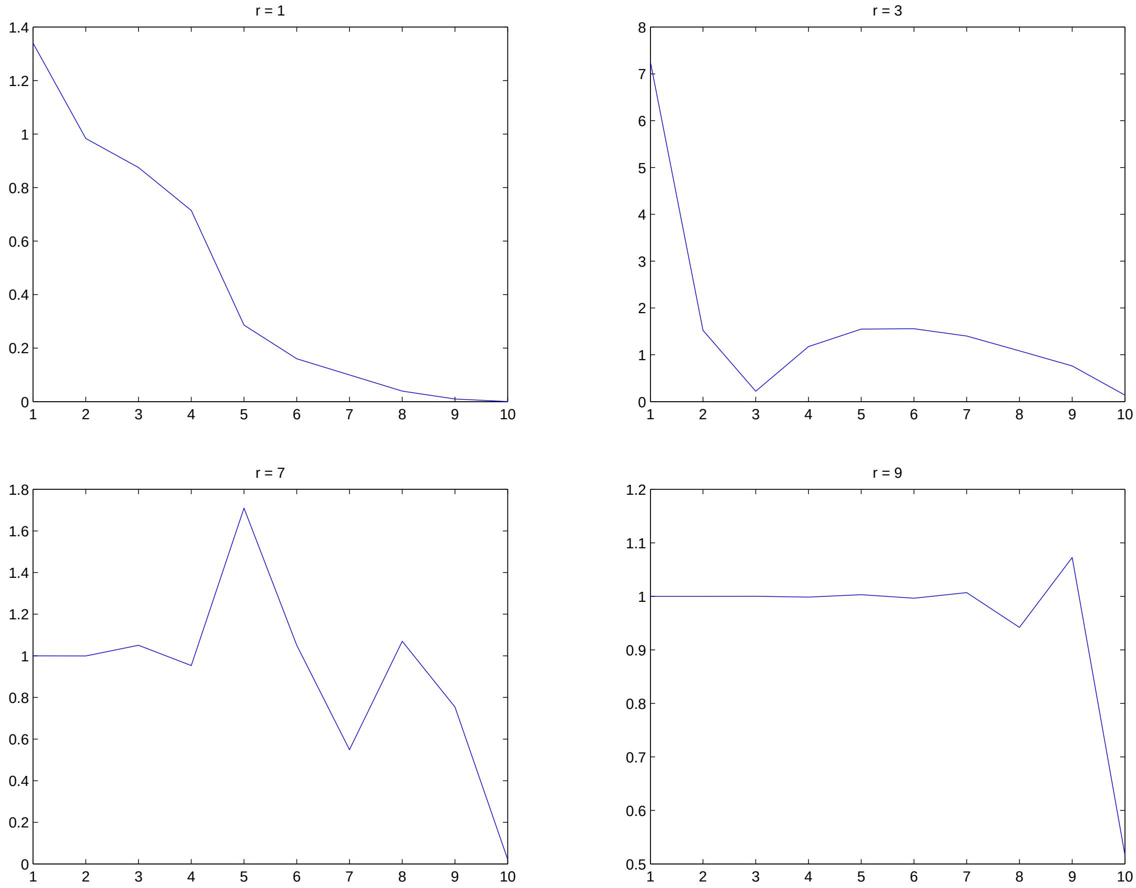

Figure 1: An illustration of the shrinkage function f_2 for different values of r. The symmetric matrix V is of dimension 10. The shrinkage is high if r is small

where $\mathbf{u}_j$ is the j^{th} eigenvector of V, $m = r - j$ with $r \geq j$, $c_j > 0$ and $\gamma_j > 1$ are constants independent of r (not considering $\hat{\sigma}$) and $T_m(.)$ is the Chebyshev polynomial of degree m. Therefore the approximation error projected on $\mathbf{u}_j$ decays as $T_m^{-1}(\gamma_j)$. The derivation of this inequality is given in the Appendix.

The Chebyshev polynomial $T_m(x)$ for $x > 1$ large m is

$$T_m(x) = \frac{1}{2}\left(e^{m\lambda} + e^{-m\lambda}\right) \simeq \frac{1}{2}e^{m\lambda}$$

where $\lambda = arccosh(x)$. Therefore $\hat{\mathbf{z}}$ converge geometrically.

4 Choosing the dimension r

Similar to the problem of determining the most appropriate dimension k when using (4), the dimension r of the Krylov subspace $K_r(V, \mathbf{y})$ plays a crucial role in finding the accurate approximation of the major eigenvectors and eigenvalues of V. This number r also corresponds to the number of steps in the algorithm described above. It can be chosen beforehand using a model selection criterion Seghouane & Bekara (2004), Seghouane & Cichocki (2007) or determined by a convergence rule tested at the end of each iteration of the algorithm. In this later case, the whitened vector is computed at the end of each step i of the algorithm.

Different convergence rule can be used, in this letter the relative norm of the difference between two consecutive approximation

$$\tau = \frac{||\hat{\mathbf{z}}_{i+1} - \hat{\mathbf{z}}_i||_2}{||\hat{\mathbf{z}}_i||_2}$$

is used as a convergence criterion. The algorithm above is stopped when this is below a certain threshold.

5 Application to fMRI Time Series Whitening

The data used to assess the performance of the proposed prewhitening method was generated from an fMRI experiment performed to investigate an event related right finger tapping task.
A 3.0 T functional MRI system was used to acquire the whole brain BOLD/EPI images. Each acquisition consisted of 35 contiguous slices ($64 \times 64 \times 64$ $3.44mm \times 3.44mm \times 4mm$ voxels). The data was recorded for a total of 650 s with TR $= 2$ s. First 30 dummy scans were discarded. After the first 30 s of rest, the task and resting period activity consisted of 14 s window was repeated 40 times followed by an additional 30 s of rest. The task period consisted of 2 s of right finger tapping. For the resting period that comes after the task, the interstimulus interval (ISI) ranged between 4 and 20 s with an average ISI period of 12 s Lee & Tak & Ye (2011).
Four data sets of size $p = 1000$, $p = 5000$, $p = 10000$ and $p = 40000$ voxels were constructed from both activated and non activated voxels. These data sets were whitened using the proposed algorithm and the truncated SVD approach implemented in the GIFT software Calhoun & Adali & Pearlson & Pekar (2001). The Q-Q plots (with Mahalanobis distance) were used to compare the underlying distribution of each of the whitened data sets with the standard normal distribution. As it can be seen on figure 1, the Q-Q plots approximately follow the line $y = x$ for whitened data sets obtained with both methods (second and third line of subfigures). This indicates that the underlying distribution of the whitened data sets on the vertical axis are identical to the standard normal. On the other hand, when looking at the computational time, table 1 indicates that the proposed whitening method is computationally more efficient than the classical approach of truncated ED or SVD.

6 Conclusion

This letter introduces a computationally efficient method for whitening high dimensional fMRI data sets based on Krylov subspaces. The gain in the computational cost of the method is

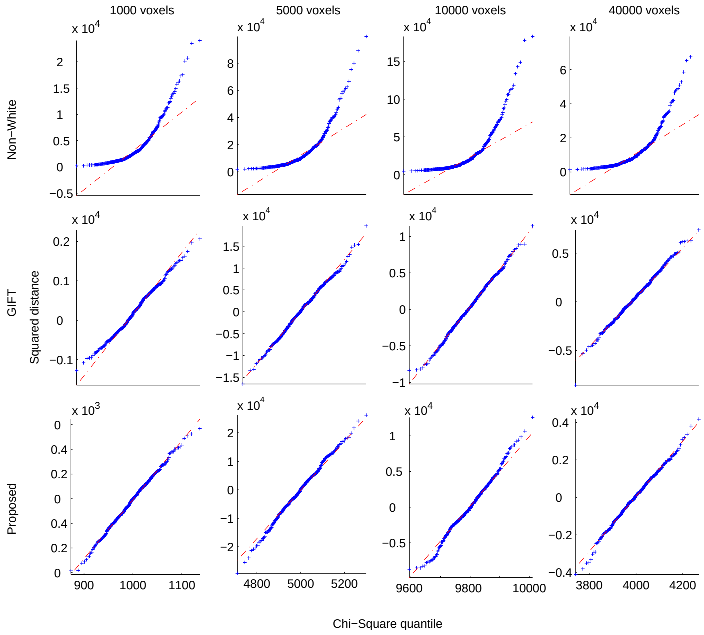

Figure 2: *QQ plots of the whitened data sets obtained from the two different methods versus the χ_p^2 distribution.*

achieved by avoiding the use of the ED of V which has a computational cost of $O(p^3)$. Instead, a Lanczos based approach is used to accurately approximate the major eigenvectors and eigenvalues of V. This has the advantage of avoiding the computation of the eigenvectors and eigenvalues which are not associated with the intrinsic dimension of the data leading to a reduced computational cost of $O(rp^2)$, $r \ll p$. As illustrated on an experiment of whitening high dimensional fMRI data, this comes without sacrificing accuracy.

No. of voxels	1000	5000	10000	40000
GIFT	4.62	10.07	17.61	64.25
Proposed	0.10	0.86	1.64	8.13

Table 1: *Computation time in seconds for the three data sets whitening obtained for the method implemented in GIFT software and the proposed method.*

Appendix

The bound on the angle between the j^{th} eigenvector $\mathbf{u}_j$ of V and the Krylov subspace defined by Q_r is given by Saad (1980)

$$||(I - Q_r Q_r^\top)\mathbf{u}_j|| \leq c_j T_m^{-1}(\gamma_j).$$

Applying the above inequality to (22)

$$\begin{aligned}
|\langle(\mathbf{z} - \hat{\mathbf{z}}), \mathbf{u}_j\rangle| &= |\langle(I - Q_r Q_r^\top)\mathbf{z}, \mathbf{u}_j\rangle| \\
&= |\langle(I - Q_r Q_r^\top)\mathbf{u}_j, \mathbf{z}\rangle| \\
&\leq ||(I - Q_r Q_r^\top)\mathbf{u}_j||_2 ||\mathbf{z}||_2 \\
&\leq c_j ||\mathbf{z}||_2 T_m^{-1}(\gamma_j).
\end{aligned}$$

References

F. G. Ashby (2011). *Statistical Analysis of fMRI data*. The MIT Press.

N. Lazar (2010). *The Statistical Analysis of Functional MRI data*. Springer.

B. Ardekani, J. Kershaw, K. Kashikura & I. Kanno (1999). 'Activation detection in functional MRI using subspace modeling and maximum likelihood estimation'. *IEEE Transactions on Medical Imaging*, **18**(2), 101–114.

A. K. Seghouane, & A. Shah (2012). 'HRF estimation in fMRI data with unknown drift matrix by iterative minimization of the Kullback-Leibler divergence'. *IEEE Transactions on Medical Imaging*, **31**(2), 192–206.

C. R. Genovese, N. A. Lazar, & T. Nichols. (2002). 'Thresholding of statistical Maps in functional neuroimaging using the False Discovery Rate'. *Neuroimage*, **15**(4), 870–878.

F. Esposito, et al. (2002). 'Spatial independent component analysis of functional MRI time series: To what extent do results depend on the alogrithm used?'. *Human Brain Mapping*, **16**(3), 146–157.

V. D. Calhoun, T. Adali, G. D. Pearlson, & J. J. Pekar. (2001). 'Spatial and Temporal independent component analysis of functional MRI data containing a pair ot task-related waveforms'. *Human Brain Mapping*, **13**(1), 43–53.

T. Adali, & V. D. Calhoun (2006). 'Unmixing fMRI with independent component analysis'. *IEEE Engineering in Medicine and Biology Magazine*, **25**(2), 79–90.

I. T. Jolliffe (2002). *Principal Component Analysis*. Springer.

A. Hyvarien, J. Karhunen, & E. Oja. (2001). *Independent Component Analysis*. Wiley.

G. H. Golub & C. f. Van Loan (1996). *Matrix Computations*. John Hopkins.

Y. Saad (1992). *Numerical Methods for Large Eigenvalues Problems*. Halstead Press, New York.

Y. C. Eldar & A. V. Oppenheim (2005). 'MMSE whitening and subspace whitening'. *IEEE Transactions on Information Theory*, **49**(7), 1846–1851.

A. K. Seghouane & M. Bekara (2004). 'A small sample model selection criterion based on Kullback's symmetric divergence'. *IEEE Transactions on Signal Processing*, **52**(12), 3314–3323.

Y. Saad (1980). 'On the rates of convergence of the Lanczos and block-Lanczos methods'. *SIAM Journal of Numerical Analysis*, **17**(5), 687–706.

V. D. Calhoun, T. Adali, G. D. Pearlson, & J. J. Pekar. (2001). 'A method for making group inferences from functional MRI data using independent component analysis'. *Human Brain Mapping*, **43**(3), 140–151.

C. Buchel & K. Friston (1997). 'Modulation of connectivity in visual pathways by attention: Cor-tical interactions evaluated with structural equation modelling and fMRI'. *Cerebral Cortex*, **7**(8), 768–778.

A. van der Sluis & H. A. van der Vorst (1997). 'The rate of convergence of conjugate gradient'. *Numerische Mathematik*, **48**(5), 543–560.

J. Chen & Y. Saad (2009). 'Lanczos vectors versus singular vectors for effective dimension reduction'. *IEEE Transactions on Knowldge and Data Engineering*, **21**(8), 1091–1103.

A. K. Seghouane & A. Cichocki (2007). 'Bayesian estimation of the number of principal components'. *Signal Processing*, **87**(3), 562–568.

K. Lee, S. Tak, & J. C. Ye (2011). 'A data-driven sparse GLM for fMRI analysis using sparse dictionary learning with MDL criterion'. *IEEE Transactions on Medical Imaging*, **30**(5), 1076–1089.