Empirical evidence of the linear nature of magnetoencephalograms

Antti Honkela, Tomas Östman and Ricardo Vigário

Neural Networks Research Centre, Helsinki University of Technology
P.O. Box 5400, FI-02015 TKK, Finland
{antti.honkela,tomas.ostman,ricardo.vigario}@tkk.fi

Abstract. Over recent years many algorithms have been used for the analysis of electro- and magnetoencephalograms, assuming a linear model for the mixing of cortical activity at the sensor plane. Such linearity can be theoretically justified, through the Maxwell equations. In this paper we exploit the adaptive and modular nature of the variational Bayesian hierarchical nonlinear factor analysis to give empirical evidence of linearity, as well as to estimate the intrinsic dimension of the generative source space.

1 Introduction

Magnetoencephalography (MEG) is a non-invasive brain imaging technique related to the electroencephalography (EEG). It is sensitive to the net magnetic flux arising from the post-synaptic currents of thousands of neurons, acting synchronously. The Maxwell equations, ruling over such electromagnetic phenomena, when applied to the various head tissue types (brain, cerebral spinal fluid, skull, scalp, ...), suggest a quasi-instantaneous linear model for the combination of cortical activity at the measuring level [1, 2].

At first as a simple first order approximation, then basing oneself on the aforementioned theoretical grounds for linearity, recent years have seen many successful linear blind source separation (BSS) approaches to the analysis of biomedical signals (c.f., [2–6]). We now give further experimental validation to the belief of a linear generative nature for MEG.

In the present study, we use the modular and adaptive nature of variational Bayesian hierarchical nonlinear factor analysis (HNFA) to show direct empirical evidence of linearity. Furthermore, we see how this algorithm is capable of estimating the intrinsic dimension of the generative space. Such estimate is crucial, e.g., when making sure we avoid overlearning phenomena [7]. In line with consistency studies of independent component analysis (ICA, c.f. [8–10]), we show also how to identify the percentage of signal and noise in such space.

2 Hierarchical nonlinear factor analysis by variational Bayesian learning

In this work, the MEG recordings are analysed with a nonlinear blind source separation method. The method is based on postprocessing the results of a nonlinear factor analysis (NFA) method with linear ICA to obtain nonlinear

independent components. NFA is a nonlinear generalisation of linear factor analysis. The generative model for observations $\mathbf{x}(t)$ in NFA is

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t)) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{s}(t)$ denotes the Gaussian sources or factors, $\mathbf{f}$ is a nonlinear mapping from sources to observations and $\mathbf{n}(t)$ is Gaussian noise. An algorithm for solving the NFA problem using a multilayer perceptron (MLP) network to model $\mathbf{f}$ is presented in [11]. The method is based on variational Bayesian learning and works well for small problems, but the computational complexity is quadratic with respect to the number of the sources and the approximately 60 sources needed to model the MEG data would increase the computational burden too much. To avoid this, a computationally simpler hierarchical NFA (HNFA) method [12] based on the variational Bayesian building block framework [13] is used instead. The HNFA method was used for reconstructing missing data in several well-known benchmark data sets in [14]. It performed clearly better than linear models, showing that the method can effectively learn a nonlinear representation of the data.

2.1 HNFA model

The model for the nonlinearity used in HNFA is layered in a way resembling MLP and radial basis function (RBF) networks. In order to attain the smaller computational complexity, the hidden nodes of the network have to be taken as additional latent variables, hence the name hierarchical NFA. The probabilistic model for the observations $\mathbf{x}(t)$ can now be written as

$$\mathbf{h}(t) \sim N(\mathbf{A}\mathbf{s}(t) + \mathbf{a}, \boldsymbol{\Sigma}_h) \tag{2}$$
$$\mathbf{x}(t) \sim N(\mathbf{C}\mathbf{s}(t) + \mathbf{B}\boldsymbol{\phi}(\mathbf{h}(t)) + \mathbf{b}, \boldsymbol{\Sigma}_x), \tag{3}$$

where $\mathbf{h}(t)$ are the additional latent variables at the hidden nodes. $\boldsymbol{\Sigma}_h$ and $\boldsymbol{\Sigma}_x$ are the noise covariance matrices of the hidden nodes and the observations, respectively, $\boldsymbol{\phi}$ is an activation function acting component-wise on its inputs and $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{a}, \mathbf{b}$ are weight matrices and bias vectors. For the sake of computational simplicity, the nonlinear activation function is $\phi(y) = \exp(-y^2)$. The structure of the model is illustrated in Fig. 1.

2.2 Variational Bayesian learning

The HNFA model is learned using a variational Bayesian method called ensemble learning [15]. It is based on approximating the posterior distribution $p(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H}|\boldsymbol{X})$ of the sources $\boldsymbol{S} = \{\mathbf{s}(t)\}$, parameters $\boldsymbol{\theta}$, and additional latent variables $\boldsymbol{H} = \{\mathbf{h}(t)\}$ with a simpler factorial approximation $q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H}) = \prod_{i,t} q(s_i(t)) \prod_{j,t} q(h_j(t)) \prod_k q(\theta_k)$. Here $\boldsymbol{X} = \{\mathbf{x}(t)\}$ is the data set. The approximation is fitted by minimising the cost

$$\mathcal{C} = \left\langle \log \frac{q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H})}{p(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{X})} \right\rangle = D_{KL}(q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H})||p(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H}|\boldsymbol{X})) - \log p(\boldsymbol{X}) \tag{4}$$
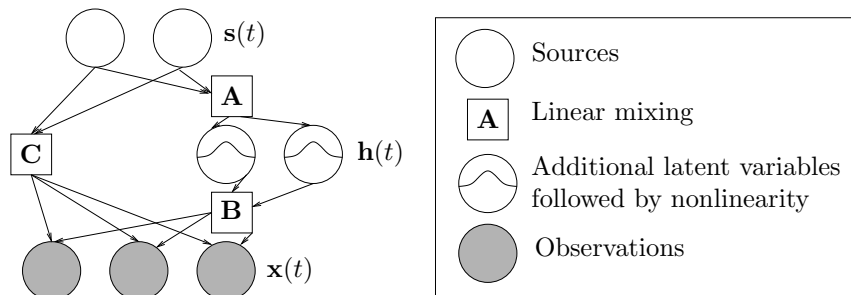
PSfrag replacements

Fig. 1: An illustration of the HNFA model. Square nodes correspond to weight matrices and round nodes to the (latent) variables.

where $\langle \cdot \rangle$ denotes expectation over $q(\boldsymbol{S}, \boldsymbol{\theta}, \boldsymbol{H})$ and $D_{KL}(q||p)$ is the Kullback-Leibler divergence between q and p. As the Kullback-Leibler divergence is always non-negative, $\mathcal{C}$ yields an upper bound for $-\log p(\boldsymbol{X})$ and thus a lower bound for the evidence $p(\boldsymbol{X})$. This allows the cost function values to be used as a criterion for model comparison and pruning of unused parts of the model [16].

2.3 Learning the model

The HNFA model is learned by minimising the cost (4) with a variational EM algorithm that updates one variable at a time while keeping the others fixed. The model structure is constructed by starting with a linear mapping from sources $\mathbf{s}(t)$ to the data $\mathbf{x}(t)$ and gradually building the nonlinearity by adding the hidden nodes $h_i(t)$. The incoming weights of the added nodes are initialised randomly. During the first addition all the generated candidates are added but later only those showing the greatest estimated decrease in the cost are selected because random candidates are unlikely to significantly decrease the cost. After adding new hidden nodes the learning algorithm is run for 30 iterations before applying pruning to remove unsuccessful additions and other redundant parts. In addition to complete hidden nodes, the individual elements of the weight matrices are also pruned and added in a similar manner. For more details on the procedure, see [12].

3 Experimental evidence

To study the linearity and the intrinsic dimensionality of the generative space of MEG recordings, the HNFA method was applied to measurements containing strong artifacts. This data was first introduced in [17], and has since constituted a good benchmarking set for the evaluation of various algorithms.

The data consisted of 122-channel whole-scalp MEG recordings, collected at 61 locations over the scalp. It contained controlled artifacts, such as eye blinks and saccades, cardiac cycle, muscle contraction through biting, as well as the contamination from a digital watch, placed inside the shielded room, one meter

away from the measuring device. The sampling frequency was 297 Hz, and the data was bandpass-filtered between 0.75 Hz and 45 Hz, and further downsampled by a factor of 2. We took 10000 samples, corresponding to around 68 seconds.

The HNFA method was tested with 48, 56, 64 and 72 sources with multiple random initialisations. The attained cost function values are shown in left panel of Fig. 2. The figure shows a clear minimum in the cost around 64 sources indicating optimal size of the model.
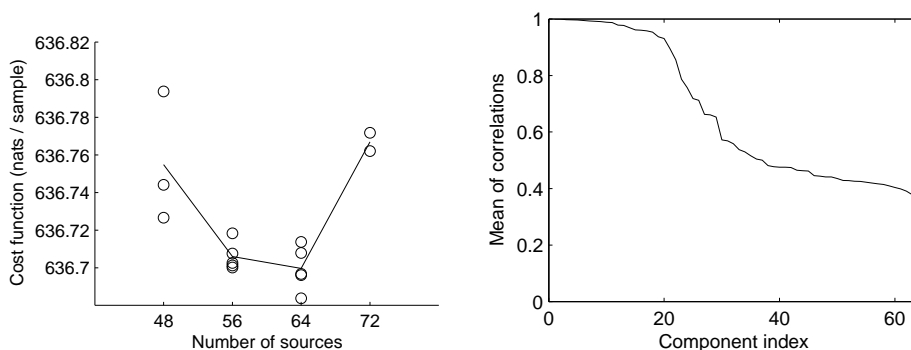

Fig. 2: Left panel: The value of the cost function as a function of the number of sources. The circles show results of individual experiments while the solid line shows the mean of the results. Right panel: Mean correlations between the components in different simulations, sorted by correlation.

All of the extracted 64 sources are not very meaningful but rather seem to be associated with more complex noise processes affecting several measurement channels. These additional signals are close to Gaussian and hence cannot be estimated reliably. This can be utilised to filter them out by running the algorithm multiple times and correlating the results of different simulations. The means of such correlations are illustrated in the right panel of Fig. 2. The most reliable components form a group of approximately 20 easily interpretable components. Selected examples of these are shown as eight topmost sources in Fig. 3 together with two less consistent noise-like components. In agreement with [17], two muscle artifacts are seen, followed by vertical and horizontal eye activity. The next two components, clearly periodic, contain contaminations to the MEG by the cardiac cycle and a digital watch, respectively. Components 7 and 8 are harder to explain physiologically, yet exhibit interesting temporal structure.

The variational Bayesian learning procedure used in HNFA allows pruning unused parts of the model. This feature is used heavily in the learning process as the model starts linear and the nonlinearity is built by gradually adding the hidden nodes h_i. In case of MEG data, practically all of the hidden nodes are pruned out. In some cases a single hidden node may remain, but the cost function value in these cases is higher than in purely linear models. This shows that the optimal HNFA model has a linear mixing mapping.
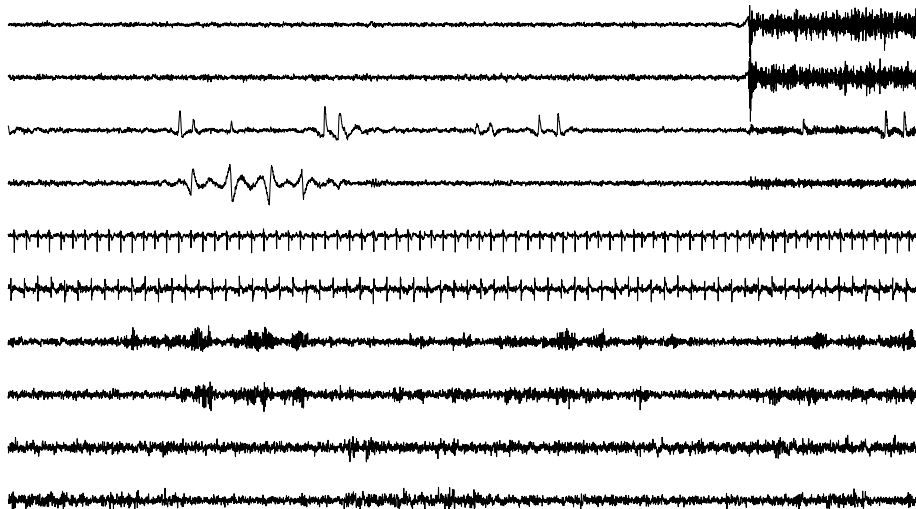
Fig. 3: Examples of the 64 sources extracted from the data.

4   Conclusions

The variational-Bayesian-based HNFA has shown in the past to be a useful method to learn nonlinear representations of data. Yet, due to its modular nature, we show here that it is also capable of pruning out that nonlinearity, when dealing with intrinsically linear data sets. This result is far from trivial, and suggests a degree of robustness against overlearning.

In a practical example of the pruning properties of HNFA, we have shown empirical evidence for the theoretically expected linear generative nature of MEG recordings. This result is important, as it validates the use of linearity in designing algorithms for processing such data.

Finally, we have been able to determine the intrinsic dimension of the generative space of this MEG data. It was done by evaluating a cost function which penalises the complexity of the suggested models, whilst helping to approximate the posterior distribution of the sources and parameters.

To distinguish between signal and noise dimensions in the estimated generative space, we have used a consistency approach. This assumes that the noise forms a close to Gaussian subspace. For each run of the algorithm, this subspace can be spanned by a different set of vectors, showing very low consistency for each individual direction. On the other hand, meaningful components are expected to be stable, hence their estimation is somewhat constant across all runs. This method reached a value of 20-25 generative signals, which is in agreement with earlier results for the same data set.

The variational approximation may in some situations favour simpler linear HNFA models over nonlinear ones. Yet, HNFA has been able to correctly identify a nonlinear generative process in several examples [14]. The results reported here

evidence clearly that MEG is a mostly linear mixture of underlying neuronal electromagnetic activity.

References

[1] M. Hämäläinen, R. Hari, R. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography—theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993.

[2] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.*, 47:589–593, 2000.

[3] S. Makeig, A. Bell, T.-P. Jung, and T. Sejnowski. Independent component analysis of electroencephalographic data. In *Advances in Neural Information Processing Systems 8*, pp. 145–151. MIT Press, 1996.

[4] R. Vigário. Extraction of ocular artifacts from EEG using independent component analysis. *Electroenceph. Clin. Neurophysiol.*, 103:395–404, 1997.

[5] T. Jung, S. Makeig, T. Lee, M. McKeown, G. Brown, A. Bell, and T. Sejnowski. Independent component analysis of biomedical signals. In *Proc. 2nd Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pp. 633–644, Helsinki, Finland, 2000.

[6] A. Tang, B. Pearlmutter, N. Malaszenko, D. Phung, and B. Reeb. Independent components of magnetoencephalography: Localization. *Neural Computat.*, 14:1827–1858, 2002.

[7] J. Särelä and R. Vigário. Overlearning in marginal distribution-based ICA: analysis and solutions. *Journal of Machine Learning Research*, 4:1447–1469, 2003.

[8] F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. Resampling approach to estimate the stability of one-dimensional or multidimensional independent components. *IEEE Trans. on Biomed. Eng.*, 49:1514–1525, 2002.

[9] J. Himberg, A. Hyvärinen, and F. Esposito. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22:1214–1222, 2004.

[10] J. Ylipaavalniemi and R. Vigário. Analysis of auditory fMRI recordings via ICA: A study on consistency. In *Proc. 2004 IEEE Int. Joint Conf. on Neural Networks (IJCNN'2004)*, Budapest, Hungary, 2004.

[11] H. Lappalainen and A. Honkela. Bayesian nonlinear independent component analysis by multi-layer perceptrons. In M. Girolami, ed., *Advances in Independent Component Analysis*, pp. 93–121. Springer-Verlag, Berlin, 2000.

[12] H. Valpola, T. Östman, and J. Karhunen. Nonlinear independent factor analysis by hierarchical models. In *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pp. 257–262, Nara, Japan, 2003.

[13] H. Valpola, T. Raiko, and J. Karhunen. Building blocks for hierarchical latent variable models. In *Proc. 3rd Int. Conf. on Independent Component Analysis and Signal Separation (ICA2001)*, pp. 710–715, San Diego, USA, 2001.

[14] T. Raiko, H. Valpola, T. Östman, and J. Karhunen. Missing values in hierarchical nonlinear factor analysis. In *Proc. Int. Conf. on Artificial Neural Networks and Neural Information Processing - ICANN/ICONIP 2003*, pp. 185–189, Istanbul, Turkey, 2003.

[15] G. E. Hinton and D. van Camp. Keeping neural networks simple by minimizing the description length of the weights. In *Proc. 6th Ann. ACM Conf. on Computational Learning Theory*, pp. 5–13, Santa Cruz, CA, USA, 1993.

[16] A. Honkela and H. Valpola. Variational learning and bits-back coding: an information-theoretic view to Bayesian learning. *IEEE Trans. Neural Networks*, 15(4):800–810, 2004.

[17] R. Vigário, V. Jousmäki, M. Hämäläinen, R. Hari, and E. Oja. Independent component analysis for identification of artifacts in magnetoencephalographics recordings. In *Advances in Neural Information Processing Systems 10*, pp. 229–235. MIT Press, 1998.