NeuroImage 49 (2010) 641–655

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Robust Bayesian estimation of the location, orientation, and time course of multiple correlated neural sources using MEG

David P. Wipf [a], Julia P. Owen [a], Hagai T. Attias [b], Kensuke Sekihara [c], Srikantan S. Nagarajan [a,*]

[a] Department of Radiology and Biomedical Imaging, University of California, San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143, USA
[b] Goldenmetallic Inc., San Francisco, USA
[c] Tokyo Metropolitan University, Tokyo, Japan

ARTICLE INFO

Article history:
Received 21 January 2009
Revised 11 June 2009
Accepted 20 June 2009
Available online 10 July 2009

ABSTRACT

The synchronous brain activity measured via MEG (or EEG) can be interpreted as arising from a collection (possibly large) of current dipoles or sources located throughout the cortex. Estimating the number, location, and time course of these sources remains a challenging task, one that is significantly compounded by the effects of source correlations and unknown orientations and by the presence of interference from spontaneous brain activity, sensor noise, and other artifacts. This paper derives an empirical Bayesian method for addressing each of these issues in a principled fashion. The resulting algorithm guarantees descent of a cost function uniquely designed to handle unknown orientations and arbitrary correlations. Robust interference suppression is also easily incorporated. In a restricted setting, the proposed method is shown to produce theoretically zero reconstruction error estimating multiple dipoles even in the presence of strong correlations and unknown orientations, unlike a variety of existing Bayesian localization methods or common signal processing techniques such as beamforming and sLORETA. Empirical results on both simulated and real data sets verify the efficacy of this approach.

Introduction

Magnetoencephalography (MEG) and related electroencephalography (EEG) use an array of sensors to take electromagnetic field (or voltage) measurements from on or near the scalp surface with excellent temporal resolution. In both MEG and EEG, the observed field can in many cases be explained by synchronous, compact current sources located within the brain. Although useful for research and clinical purposes, accurately determining the spatial distribution of these unknown sources is a challenging inverse problem. The relevant estimation problem can be posed as follows: the measured electromagnetic signal is $B \in \mathbb{R}^{d_b \times d_t}$, where d_b equals the number of sensors and d_t is the number of time points at which measurements are made. Each unknown source $S_i \in \mathbb{R}^{d_c \times d_t}$ is a d_c-dimensional neural current dipole, at d_t time points, projecting from the i-th (discretized) voxel or candidate location distributed throughout the brain. These candidate locations can be obtained by segmenting a structural MR scan of a human subject and tesselating the brain volume with a set of vertices. B and each S_i are related by the likelihood model

$$B = \sum_{i=1}^{d_s} L_i S_i + \varepsilon, \tag{1}$$

where d_s is the number of voxels under consideration and $L_i \in \mathbb{R}^{d_b \times d_c}$ is the so-called lead-field matrix for the i-th voxel. The k-th column of L_i represents the signal vector that would be observed at the scalp given a unit current source/dipole at the i-th vertex with a fixed orientation in the k-th direction. It is common to assume $d_c = 2$ (for MEG) or $d_c = 3$ (for EEG), which allows flexible source orientations to be estimated in 2D or 3D space. Multiple methods based on the physical properties of the brain and Maxwell's equations are available for the computation of each L_i (Sarvas, 1987). Finally, ε is a noise-plus-interference term where we assume, for simplicity, that the columns are drawn independently from $N(0,\Sigma_\varepsilon)$. However, temporal correlations can easily be incorporated if desired using a simple transformation outlined in Friston et al. (2008) or using the spatio-temporal framework introduced in Bolstad et al. (2009).

To obtain reasonable spatial resolution, the number of candidate source locations will necessarily be much larger than the number of sensors ($d_s \gg d_b$). The salient inverse problem then becomes the ill-posed estimation of regions with significant brain activity, which are reflected by voxels i such that $||S_i|| > 0$; we refer to these as active dipoles or sources. Because the inverse model is severely underdetermined (the mapping from source activity configuration $S \triangleq \left[S_1^T, \ldots, S_{d_s}^T\right]^T$ to sensor measurement B is many to one), all efforts at source reconstruction are heavily dependent on prior assumptions, which in a Bayesian framework are embedded in the distribution $p(S)$. Such a prior is often considered to be fixed and known, as in the case of minimum ℓ_2-norm estimation (MNE) (Baillet et al., 2001), minimum current estimation (MCE) (Uutela et al., 1999), minimum variance adaptive beamforming (MVAB) (Sekihara and Nagarajan, 2008), FOCUSS (Gorodnitsky et al., 1995), and sLORETA (Pascual-Marqui, 2002). Alternatively, empirical Bayesian approaches have

* Corresponding author. Fax: +1 415 502 4302.
E-mail address: sri@radiology.ucsf.edu (S.S. Nagarajan).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2009.06.083

been proposed that attempt a form of model selection by using the data, whether implicitly or explicitly, to guide the search for an appropriate prior. Examples include a rich variety of variational Bayesian methods and hierarchical covariance component models (Friston et al., 2008; Mattout et al., 2006; Phillips et al., 2005; Sahani and Nagarajan, 2004; Sato et al., 2004; Wipf et al., 2007; Zumer et al., 2007). These hierarchical models can also be handled by replacing variational inference and related procedures with Markov-Chain Monte Carlo sampling (Nummenmaa et al., 2007). In general, these approaches differ in the types of covariance components that are assumed, the cost functions and approximations used to estimate unknown parameters, and in how posterior distributions are invoked to affect source localization. While advantageous in many respects, none of these methods are explicitly designed to handle complex, correlated source configurations with unknown orientation in the presence of background interference (e.g., spontaneous brain activity, sensor noise, etc.).

There are two types of correlations that can potentially disrupt the source localization process. First, there are correlations within dipole components (meaning the individual rows of S_i are correlated), which always exists to a high degree in real data (i.e., $d_c>1$). For example, dipoles with a fixed unknown orientation will have a correlation coefficient of 1.0; for rotating or wobbling dipoles this value will typically be smaller. Secondly, there are correlations between different dipoles that are simultaneously active (meaning rows of S_i are correlated with rows of S_j for some voxels $i\neq j$). These correlations are more application specific and may or may not exist. The larger the number of active sources, the greater the chance that both types of correlations can disrupt the estimation process. This issue can be problematic for two reasons. First, and most importantly, failure to accurately account for unknown orientations or correlations can severely disrupt the localization process, leading to a very misleading impression of which brain areas are active. Secondly, the orientations and correlations themselves may have clinical significance, although this will not be our focus herein.

In this paper, we present an alternative empirical Bayesian scheme that attempts to improve upon existing methods in terms of source reconstruction accuracy and/or computational robustness and efficiency. The Modeling assumptions section presents the basic generative model which underlies the proposed method and describes the associated inference problem. This model is designed to estimate the number and location of a small (sparse) set of flexible dipoles that adequately explain the observed sensor data. The Algorithm derivation section derives a robust algorithm, which we call *Champagne*, for estimating the sources using this model and proves that each iteration is guaranteed to reduce the associated cost function. It also describes how interference suppression is naturally incorporated. The Conditions for perfect reconstruction section then provides a theoretical analysis of conditions under which perfect reconstruction of arbitrary, correlated sources with unknown orientation is possible, demonstrating that the proposed method has substantial advantages over existing approaches. Finally, the Empirical evaluation section contains experimental results using our algorithm on both simulated and real data, followed by brief conclusions in the Conclusion section.

Modeling assumptions

To begin we invoke the noise model from Eq. (1), which fully defines the assumed likelihood

$$p(B|S)\propto\exp\left(-\frac{1}{2}\left\|B-\sum_{i=1}^{d_s}L_iS_i\right\|^2_{\Sigma_\varepsilon^{-1}}\right), \tag{2}$$

where $\|X\|_W$ denotes the weighted matrix norm $\sqrt{\mathrm{trace}[X^TWX]}$. The unknown noise covariance Σ_ε will be estimated from the data using a variational Bayesian factor analysis (VBFA) model as discussed in the Learning the interference Σ_ε section; for now we will consider that it is fixed and known. Next we adopt the following source prior for S:

$$p(S|\Gamma)\propto\exp\left(-\frac{1}{2}\mathrm{trace}\left[\sum_{i=1}^{d_s}S_i^T\Gamma_i^{-1}S_i\right]\right). \tag{3}$$

This is equivalent to applying independently, at each time point, a zero-mean Gaussian distribution with covariance Γ_i to each source S_i. We define Γ to be the $d_sd_c\times d_sd_c$ block-diagonal matrix formed by ordering each Γ_i along the diagonal of an otherwise zero-valued matrix. This implies, equivalently, that $p(S|\Gamma)\propto\exp\left(-\frac{1}{2}\mathrm{trace}\left[S^T\Gamma^{-1}S\right]\right)$.

If Γ were somehow known, then the conditional distribution $p(S|B,\Gamma)\propto p(B|S)p(S|\Gamma)$ is a fully specified Gaussian distribution with mean and covariance given by

$$\begin{aligned}&E_{p(S|B,\Gamma)}[S]=\Gamma L^T(\Sigma_\varepsilon+L\Gamma L^T)^{-1}B\\&\mathrm{Cov}_{p(\boldsymbol{s}_j|B,\Gamma)}\left[\boldsymbol{s}_j\right]=\Gamma-\Gamma L^T\left(\Sigma_\varepsilon+L\Gamma L^T\right)^{-1}L\Gamma,\quad\forall j,\end{aligned} \tag{4}$$

where $L\triangleq[L_1,\ldots,L_{d_s}]$ and $\boldsymbol{s}_j$ denotes the j-th column of S (i.e., the sources at the j-th time point) and individual columns are uncorrelated. However, since Γ is actually not known, a suitable approximation $\hat{\Gamma}\approx\Gamma$ must first be found. One principled way to accomplish this is to integrate out the sources S and then maximize

$$\begin{aligned}&p(B|\Gamma)=\int p(B|S)p(S|\Gamma)dS\propto|\Sigma_b|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}B^T\Sigma_b^{-1}B\right),\\&\Sigma_b\triangleq\Sigma_\varepsilon+L\Gamma L^T.\end{aligned} \tag{5}$$

This is equivalent to minimizing the cost function

$$\mathcal{L}(\Gamma)\triangleq-2\log p(B|\Gamma)\equiv\mathrm{trace}\left[C_b\Sigma_b^{-1}\right]+\log|\Sigma_b|, \tag{6}$$

where $C_b\triangleq d_t^{-1}BB^T$ is the empirical covariance. This process is sometimes referred to as type-II maximum likelihood, evidence maximization, or empirical Bayes (Berger, 1985).

The first term of Eq. (6) is a measure of the dissimilarity between the empirical data covariance C_b and the model data covariance Σ_b; in general, this factor encourages Γ to be large because it is convex and nonincreasing in Γ (in a simplified scalar case, this is akin to minimizing $1/x$ with respect to x, which of course naturally favors x being large). The second term provides a regularizing or sparsifying effect, penalizing a measure of the volume formed by the model covariance Σ_b.[1] Since the volume of any high dimensional space is more effectively reduced by collapsing individual dimensions as close to zero as possible (as opposed to incrementally reducing all dimensions isometrically), this penalty term promotes a model covariance that is maximally degenerate (or non-spherical), which pushes elements of Γ to exactly zero (hyperparameter sparsity). This intuition is supported theoretically by the results in the Conditions for perfect reconstruction section.

Given some type-II ML estimate $\hat{\Gamma}$ computed by minimizing Eq. (6), we obtain the attendant empirical prior $p(S|\hat{\Gamma})$. To the extent that this 'learned' prior is realistic, the resulting posterior $p(S|B,\hat{\Gamma})$ quantifies regions of significant current density and point estimates for the unknown source dipoles S_i can be obtained by evaluating the posterior mean computed using Eq. (4). If a given $\hat{\Gamma}_i\to 0$ as described above, then the associated $\hat{S}_i$ computed using Eq. (4) also becomes zero. It is this pruning mechanism that naturally chooses the number of active dipoles and is consistent with the hypothesis that most regions of the brain are approximately inactive for a given task.

[1] The determinant of a matrix is equal to the product of its eigenvalues, a well-known volumetric measure.

Algorithm derivation

Given Σ_ϵ and Γ, computing the Gaussian posterior on S is straightforward as outlined above. Consequently, determining these unknown quantities is the primary estimation task. We will first derive an algorithm for computing Γ assuming Σ_ϵ is known. Later in the Learning the interference Σ_ϵ section, we will describe a powerful procedure for learning Σ_ϵ.

Learning the hyperparameters Γ

The primary objective of this section is to minimize Eq. (6) with respect to Γ. Of course one option is to treat the problem as a general nonlinear optimization task and perform gradient descent or some other generic procedure. Alternatively, several methods in the MEG literature rely, either directly or indirectly, on a form of the expectation maximization (EM) algorithm (Friston et al., 2008; Sato et al., 2004). However, these algorithms are exceedingly slow when d_s is large and they have not been extended to handle arbitrary, unknown orientations. Consequently, here we derive an optimization procedure that expands upon ideas from Sato et al. (2004) and Wipf et al. (2007), handles arbitrary/unknown dipole orientations, and converges quickly.

To begin, we note that $\mathcal{L}(\Gamma)$ only depends on the data B through the $d_b \times d_b$ sample correlation matrix C_b. Therefore, to reduce the computational burden, we replace B with a matrix $\widetilde{B} \in \mathbb{R}^{d_b \times \mathrm{rank}(B)}$ such that $\widetilde{B}\widetilde{B}^T = C_b$. This removes any per-iteration dependency on d_t, which can potentially be large, without altering the actual cost function. It also implies that, for purposes of computing Γ, the number of columns of S is reduced to match rank(B). We now re-express the cost function $\mathcal{L}(\Gamma)$ in an alternative form leading to convenient update rules and, by construction, a proof that $\mathcal{L}\left(\Gamma^{(k+1)}\right) \leq \mathcal{L}\left(\Gamma^{(k)}\right)$ at each iteration.

The procedure we will use involves constructing auxiliary functions using sets of hyperplanes; an introduction to the basic ideas can be found in Appendix A. First, the log-determinant term of $\mathcal{L}(\Gamma)$ is a concave function of Γ and so it can be expressed as a minimum over upper-bounding hyperplanes via

$$\log|\Sigma_b| = \min_{Z \geq 0} \left[\sum_{i=1}^{d_s} \mathrm{trace}\left(Z_i^T \Gamma_i\right) - h^*(Z) \right], \tag{7}$$

where $Z \triangleq \left[Z_1^T, \ldots, Z_{d_s}^T\right]^T$ is a matrix of auxiliary variables that differentiates each hyperplane and $h^*(Z)$ is the concave conjugate of $\log|\Sigma_b|$. While $h^*(Z)$ is unavailable in closed form, for our purposes below, we will never actually have to compute this function. Next, the data fit term is a concave function of Γ^{-1} and so it can also be expressed using similar methodology as

$$\mathrm{trace}\left[C_b \Sigma_b^{-1}\right] = \min_X \left[\left\| \widetilde{B} - \sum_{i=1}^{d_s} L_i X_i \right\|_{\Sigma_\epsilon^{-1}}^2 + \sum_{i=1}^{d_s} \left\| X_i \right\|_{\Gamma_i^{-1}}^2 \right], \tag{8}$$

where $X \triangleq \left[X_1^T, \ldots, X_{d_s}^T\right]^T$ is a matrix of auxiliary variables as before. Note that in this case, the implicit concave conjugate function exists in closed form.

Dropping the minimizations and combining terms from Eqs. (7) and (8) lead to the modified cost function

$$\mathcal{L}(\Gamma, X, Z) = \left\| \widetilde{B} - \sum_{i=1}^{d_s} L_i X_i \right\|_{\Sigma_\epsilon^{-1}}^2 + \sum_{i=1}^{d_s} \left[\left\| X_i \right\|_{\Gamma_i^{-1}}^2 + \mathrm{trace}\left(Z_i^T \Gamma_i\right) \right] - h^*(Z), \tag{9}$$

where by construction $\mathcal{L}(\Gamma) = \min_X \min_Z \mathcal{L}(\Gamma, X, Z)$. It is straightforward to show that if $\{\hat{\Gamma}, \hat{X}, \hat{Z}\}$ is a local (global) minimum to $\mathcal{L}(\Gamma, X, Z)$, then Γ is a local (global) minimum to $\mathcal{L}(\Gamma)$.

Since direct optimization of $\mathcal{L}(\Gamma)$ may be difficult, we can instead iteratively optimize $\mathcal{L}(\Gamma, X, Z)$ via coordinate descent over Γ, X, and Z. In each case, when two are held fixed, the third can be globally minimized in closed form. This ensures that each cycle will reduce $\mathcal{L}(\Gamma, X, Z)$, but more importantly, will reduce $\mathcal{L}(\Gamma)$ (or leave it unchanged if a fixed-point or limit cycle is reached). The associated update rules from this process are as follows.

The optimal X (with Γ and Z fixed) is just the standard weighted minimum-norm solution given by

$$X_i^{\mathrm{new}} \rightarrow \Gamma_i L_i^T \Sigma_b^{-1} \widetilde{B} \tag{10}$$

for each i. The minimizing Z equals the slope at the current Γ of $\log|\Sigma_b|$, which follows from simple geometric considerations (Boyd and Vandenberghe, 2004). As such, we have

$$Z_i^{\mathrm{new}} \rightarrow \nabla_{\Gamma_i} \log|\Sigma_b| = L_i^T \Sigma_b^{-1} L_i. \tag{11}$$

With Z and X fixed, computing the minimizing Γ is a bit more difficult because of the constraint $\Gamma_i \in H^+$ for all i, where H^+ is the set of positive-semidefinite, symmetric $d_c \times d_c$ covariance matrices. To obtain each Γ_i, we must solve

$$\Gamma_i^{\mathrm{new}} \rightarrow \arg\min_{\Gamma_i \in H^+} \left[\left\| X_i \right\|_{\Gamma_i^{-1}}^2 + \mathrm{trace}\left(Z_i^T \Gamma_i\right) \right]. \tag{12}$$

An unconstrained solution will satisfy

$$\nabla_{\Gamma_i} \mathcal{L}(\Gamma_i, X_i, Z_i) = 0, \tag{13}$$

which, after computing the necessary derivatives and re-arranging terms gives the equivalent condition

$$X_i X_i^T = \Gamma_i Z_i \Gamma_i. \tag{14}$$

There are multiple (unconstrained) solutions to this equation; we will choose the unique one that satisfies the constraint $\Gamma_i \in H^+$. This can be found using

$$\begin{aligned} X_i X_i^T &= Z_i^{-1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right) Z_i^{-1/2} \\ &= Z_i^{-1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right)^{1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right)^{1/2} Z_i^{-1/2} \\ &= \left[Z_i^{-1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right)^{1/2} Z_i^{-1/2} \right] Z_i \\ &\quad \times \left[Z_i^{-1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right)^{1/2} Z_i^{-1/2} \right]. \end{aligned} \tag{15}$$

This indicates (via simple pattern matching) the solution (or update equation)

$$\Gamma_i^{\mathrm{new}} \rightarrow Z_i^{-1/2}\left(Z_i^{1/2} X_i X_i^T Z_i^{1/2}\right)^{1/2} Z_i^{-1/2}, \tag{16}$$

which satisfies the constraint. And since we are minimizing a convex function of Γ_i (over the constraint set), we know that this is indeed a minimizing solution.

In summary then, to estimate Γ, we need simply iterate Eqs. (10), (11), and (16), and with each pass we are guaranteed to reduce (or leave unchanged) $\mathcal{L}(\Gamma)$; we refer to the resultant algorithm as *Champagne*. The per-iteration cost is linear in the number of voxels d_s so the computational cost is relatively modest (it is quadratic in d_b, and cubic in d_c, but these quantities are relatively small). The convergence rate is orders of magnitude faster than EM-based algorithms such as those in Friston et al. (2008) and Sato et al. (2004) (Empirical evaluation section contains a representative example).

Simplifications using constraints on Γ

Here we consider two constraints on the parameterization of Γ that lead to much less complex updates and provide connections with existing algorithms. First, we can require that off-diagonal terms of each Γ_i are equal to zero, i.e., the prior covariance of each source S_i is diagonal. We then restrict ourselves to learning the diagonal elements of Γ_i via simplified versions of those presented above. A second possibility is to further constrain each Γ_i to satisfy $\Gamma_i = \gamma_i I$, where γ_i is a scalar non-negative hyperparameter. In both cases, the resulting cost functions and algorithms also fall out of the framework we discuss in Wipf (2006) and Wipf et al. (2007). These variants are also similar to the covariance component estimation model used in Mattout et al. (2006) and Friston et al. (2008), albeit with a much larger number of components here.

As we will see in the Conditions for perfect reconstruction and Empirical evaluation sections, the reduced parameterization associated with these models can potentially degrade performance in some situations. Nonetheless, these approaches are very useful for comparison purposes. To distinguish all three cases, we use the designations CHAMP_S for the scalar version, CHAMP_D, for the arbitrary diagonal case, and CHAMP_M when the matrices Γ_i are unrestricted. For the special case where $d_\text{c} = 2$, this gives the following parameterizations:

$$\underset{\Gamma_i = \begin{bmatrix} \gamma_i & 0 \\ 0 & \gamma_i \end{bmatrix}}{\text{CHAMP}_\text{S}} \quad \underset{\Gamma_i = \begin{bmatrix} \gamma_{i_1} & 0 \\ 0 & \gamma_{i_2} \end{bmatrix}}{\text{CHAMP}_\text{D}} \quad \underset{\Gamma_i = \begin{bmatrix} \gamma_{i_{11}} & \gamma_{i_{12}} \\ \gamma_{i_{12}} & \gamma_{i_{22}} \end{bmatrix}}{\text{CHAMP}_\text{M}}. \tag{17}$$

Learning the interference Σ_ε

The learning procedure described in the previous section boils down to fitting a structured maximum likelihood covariance estimate $\Sigma_\text{b} = \Sigma_\varepsilon + L\Gamma L^T$ to the data covariance C_b. The idea here is that $L\Gamma L^T$ will reflect the brain signals of interest while Σ_ε will capture all interfering factors, e.g., spontaneous brain activity, sensor noise, muscle artifacts, etc. Since Σ_ε is unknown, it must somehow be estimated or otherwise accounted for. Given access to pre-stimulus data (i.e., data assumed to have no signal/sources of interest), stimulus evoked partitioned factor analysis provides a powerful means of decomposing a data covariance matrix C_b into signal and interference components. While details can be found in Nagarajan et al. (2007), the procedure computes the approximation

$$C_\text{b} \approx \Lambda + \text{EE}^T + \text{FF}^T, \tag{18}$$

where $E \in \mathbb{R}^{d_\text{b} \times d_\text{e}}$ represents a matrix of learned interference factors, Λ is a diagonal noise matrix, and $F \in \mathbb{R}^{d_\text{b} \times d_\text{f}}$ represents signal factors. Both the number of interference factors d_e and the number of signal factors d_f are learned from the data via a variational Bayesian factor analysis procedure. Using a generalized form of the expectation maximization algorithm, the method attempts to find a small number of factors that adequately explains the observed sensor data covariance during both the pre- and post-stimulus periods. The pre-stimulus is modeled with a covariance restricted to the terms $\Lambda + \text{EE}^T$, while the post-stimulus covariance, which contains the signal information FF^T we wish to localize, is expressed additively as in Eq. (18).

There are two ways to utilize the decomposition (Eq. (18)). First, we can simply set $\Sigma_\varepsilon \rightarrow \Lambda + \text{EE}^T$ and proceed as in the Learning the hyperparameters section. Alternatively, we can set $\Sigma_\varepsilon \rightarrow 0$ (or set it to a small diagonal matrix for robustness/stability) and then substitute FF^T for C_b, i.e., run the same algorithm on a de-noised signal covariance. In practice, both methods have been successful; however, a full technical discussion of the relative merits is beyond the scope of this paper.

Conditions for perfect reconstruction

Whenever a new inverse algorithm is proposed, it is often insightful to know what source configurations it can recover exactly and under what conditions. While in general this will require unrealistic assumptions such as zero noise and a perfect forward model, the underlying idea is that if only implausible source configurations can be recovered even given these strong simplifications, then perhaps the algorithm may have serious difficulties.

For example, consider the classical MNE method (Baillet et al., 2001). Here exact recovery of the true sources requires that columns of the source activity matrix S lie in the range space of the transposed lead-field, i.e., $S = L^T\mathcal{A}$, for some coefficient matrix $\mathcal{A}$. But this is a very contrived requirement, because the lead-field transpose is highly overdetermined (many more rows than columns) meaning that the true S must be constrained to a small subspace of the total possible source space, i.e., a subspace of $\mathbb{R}^{d_\text{s} d_\text{c} \times d_\text{t}}$. But this subspace has nothing to do with any source activity or neurophysiological plausibility. The lead-field, which determines this subspace, only specifies how a given source configuration maps to the sensors, it is unrelated to what the actual activity is. Therefore, MNE is effectively constraining the possible source space in an ad hoc manner even before confounding noise or forward modeling errors are introduced.

A related concept is the notion of localization bias, which is basically a way of quantifying the ability of an algorithm to accurately (or sometimes perfectly) estimate a single dipolar source. Recently it has been shown, both empirically and theoretically (Pascual-Marqui, 2002; Sekihara et al., 2005), that the MVAB and sLORETA algorithms have zero location bias given noiseless data and an ideal forward model, meaning their respective source estimates will peak exactly at the true source. Note that this is a slightly different notion than perfect reconstruction, since both methods will still (incorrectly) produce non-zero source estimates elsewhere. These ideas have been extended to include certain empirical Bayesian methods (Sato et al., 2004; Wipf et al., 2007). However, these results assume a single dipole with fixed, known orientation (i.e., $d_\text{c} = 1$), and therefore do not formally handle more complex issues that arise when multiple dipoles with unknown orientations are present. The methods from Sahani and Nagarajan (2004) and Zumer et al. (2007) also purport to address these issues, but no formal analyses are presented.

Despite its utilization of a complex, non-convex cost function $\mathcal{L}(\Gamma)$, we now demonstrate relatively general conditions whereby Champagne can exhibit perfect reconstruction. We will assume that the full lead-field L represents a sufficiently high sampling of the source space such that any active dipole component aligns with some lead-field columns (the ideal forward model assumption). Perfect reconstruction can also be shown in the continuous case, but the discrete scenario is more straightforward and of course more relevant to any practical task.

Some preliminary definitions are required to proceed. We define the empirical intra-dipole correlation matrix at the i-th voxel as $C_{ii} \triangleq \frac{1}{d_\text{t}} S_i S_i^T$; non-zero off-diagonal elements imply that correlations are present. Except in highly contrived situations, this type of correlation will always exist. The empirical inter-dipole correlation matrix between voxels i and j is $C_{ij} \triangleq \frac{1}{d_\text{t}} S_i S_j^T$; any non-zero element implies the existence of a correlation. In practice, this form of correlation may or may not be present. With regard to the lead-field L, spark is defined as the smallest number of linearly dependent columns (Donoho and Elad, 2003). By definition then, $2 \leq \text{spark}\,(L) \leq d_\text{b} + 1$. Finally, d_a denotes the number of active sources, i.e., the number of voxels whereby $||S_i|| > 0$.

Theorem 1. In the limit as $\Sigma_\varepsilon \rightarrow 0$ (high SNR) and assuming $d_\text{a} d_\text{c} < \text{spark}\,(L) - 1$, the cost function $\mathcal{L}(\Gamma)$ maintains the following two properties:

1) For arbitrary C_{ii} and C_{ij}, the unique global minimum Γ^* produces a source estimate $S^* = E_{p(S|B,\Gamma^*)}[S]$ computed using Eq. (4) that

equals the generating source matrix S, i.e., it produces a perfect source reconstruction.

2) If $C_{ij} = 0$ for all active dipoles (although C_{ii} is still arbitrary), then there are no local minima, i.e., the cost function is unimodal.

See the Appendix B for the proof. In words, this theorem says that intra-dipole correlations do not disrupt the estimation process by creating local minima, and that the global minimum always achieves a perfect source reconstruction. In contrast, inter-dipole correlations can potentially create local minima, but they do not affect the global minimum. Empirically, we will demonstrate that the algorithm derived in the Algorithm derivation section is effective at avoiding these local minima (see Empirical evaluation section). With added assumptions these results can be extended somewhat to handle the inclusion of noise.

The cost functions from Sato et al. (2004) and Wipf et al. (2007) bear the closest resemblance to $\mathcal{L}(\Gamma)$; however, neither possesses the second attribute from Theorem 1. This is a significant failing because, as mentioned previously, intra-dipole correlations are always present in each active dipole. Consequently, reconstruction errors can occur because of convergence to a local minimum. The iterative Bayesian scheme from Zumer et al. (2007), while very different in structure, also directly attempts to estimate flexible orientations and handle, to some extent, source correlations. While details are omitted for brevity, we can prove that the full model upon which this algorithm is based fails to satisfy the first property of the theorem, so the corresponding global minimum can fail to reconstruct the sources perfectly. In contrast, beamformers and sLORETA are basically linear methods with no issue of global or local minima. However, the popular sLORETA and MVAB solutions will in general display peaks that may be misaligned from the true sources for multi-component dipoles ($d_c > 1$) or when multiple dipoles ($d_a > 1$) are present, regardless of correlations (they will of course never produce perfect reconstructions of dipolar sources because of spatial blur).

So in summary, our analysis provides one level of theoretical comparison. In some sense it is relevant to questions of the sort, do we expect the underlying true current sources to lie in the subspace formed by $S = L^T\mathcal{A}$, where L^T is essentially an arbitrary overdetermined matrix in this context, or will they more likely be of the form S equals the sum of $d_a < \frac{spark(L)-1}{d_c}$ arbitrary dipoles with unconstrained orientations and locations? This latter configuration is a actually equivalent to the union of a large number of $d_a d_c$-dimensional subspaces (in fact the number is combinatorial, $\binom{d_s}{d_a}$, which represents all possible unique configurations of d_a dipoles), a far more complex, and in some sense richer, set of possible sources. Champagne is designed to recover sources resembling the latter, an approximation to 'real' source configurations that many neurophysiologists would say is both plausible and useful clinically, especially in an event-related paradigm.

Empirical evaluation

In this section we test the performance of our algorithm on both simulated and real data sets. We focus here on localization accuracy assuming strong source correlations and unknown orientations. Note that the primary purpose of the proposed algorithm is not really to estimate the actual orientations or correlations per se. It is to accurately estimate the location and power of sources confounded by the effects of unknown orientations and correlations. Consequently, accurate localization estimates implicitly indicate that these confounds have been adequately handled, hence orientation (or correlation) estimates themselves are not stressed.

Simulated data

We first conducted tests using simulated data with realistic source configurations. The brain volume was segmented into 5 mm voxels and a two orientation ($d_c = 2$) forward lead-field was calculated using a single spherical-shell model (Sarvas, 1987). The data time courses were partitioned into pre- and post-stimulus periods. In the pre-stimulus period (263 samples) there is only noise and interfering brain activity, while in the post-stimulus period (437 samples) there is the same (statistically) noise and interference factors plus source activity of interest. The pre-stimulus activity consisted of the resting-state sensor recordings collected from a human subject presumed to have spontaneous activity (i.e., non-stimulus evoked sources) and sensor noise; this activity was on-going and continued into the post-stimulus period, where the simulated source signals were added. Source time courses were seeded at locations in the brain with damped-sinusoidal signals and this voxel activity was projected to the sensors through the lead-field. The locations for the sources were chosen so that there was some minimum distance between sources and a minimum distance from the center of the head. We could adjust both the signal-to-noise-plus-interference ratio (SNIR), the correlations between the different voxel time courses (inter-dipole), and the correlations between the two orientations of the dipoles (intra-dipole) to examine the algorithm performance on unknown correlated sources and dipole orientations. For our purposes, SNIR is defined as

$$\text{SNIR} \triangleq 20 \log \frac{\|LS\|_{\mathcal{F}}}{\|\mathcal{E}\|_{\mathcal{F}}}. \tag{19}$$

To obtain aggregate data on the performance of our method on many different dipole configurations and noise levels, we ran 100 simulations of three randomly (located) seeded sources at SNIR levels of -5, 0, 5, 10 dB. The sources in these simulations always had an inter-dipole correlation coefficient and an intra-dipole correlation coefficient of 0.95. We chose to test our algorithm against five representative source localization algorithms from the literature: minimum variance adaptive beamforming (MVAB) (Sekihara and Nagarajan, 2008), two non-adaptive spatial filtering methods, sLORETA (Pascual-Marqui, 2002) and dSPM (Dale et al., 2000), and two variants of minimum current estimation (MCE) specially tailored to handle multiple time points and unconstrained dipoles (Wipf, 2006; Wipf and Nagarajan, 2009). These two methods extend standard MCE by applying a ℓ_2 norm penalty across time (an ℓ_1 norm over space and an ℓ_2 norm over time, sometimes called an $\ell_{1,2}$ norm in signal processing). In one case both dipole components are also included within the ℓ_2 penalty (MCE_1), in the other case they are treated individually (MCE_2). Similar to Champagne, both versions of MCE favor sparse/compact source reconstructions. Related algorithms are discussed in Bolstad et al. (2009) and Huang et al. (2006).

We ran the simulations using a total of eight algorithms, the five above plus the three variants of our Champagne method: CHAMP_S, CHAMP_D, and CHAMP_M (see Simplifications using constraints on Γ section for a description of these variants). In order to evaluate performance, we used two features: source localization accuracy and time-course estimation accuracy. To assess localization accuracy, we used the A' metric (Snodgrass and Corwin, 1988) and to assess the estimation of the time courses, we used the correlation coefficient between the true and estimated time courses.

When assessing the localization accuracy, it is important to take into account both the number of hits (sources that were correctly localized) and the presence of false positives or spurious localizations. A principled way to take these two features into account is the ROC (receiver-operator characteristic) method modified for brain imaging results (Darvas et al., 2004), which is a measure of hit rate (HR) versus false positive rate (FR). Specifically, we used the A' metric which is a way to approximate the area under the ROC for one HR/FR pair. We determined the HR and FR at each SNIR level and each algorithm in the following way. For each simulation, we calculated all the local peaks in the image. A local peak is defined as a voxel that is greater than its 30

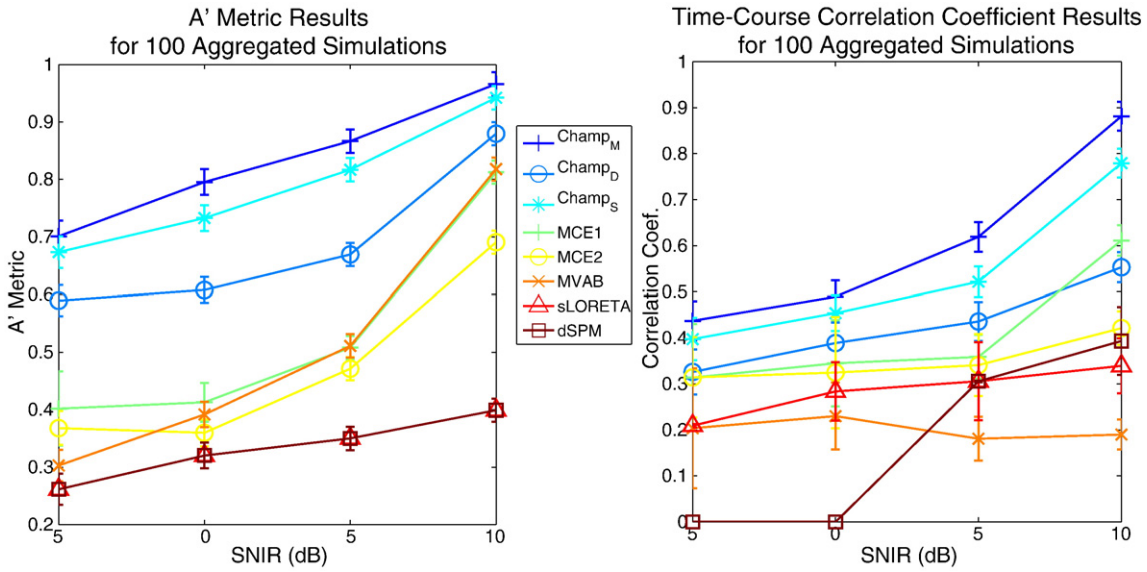

Fig. 1. Performance evaluation. Left: Aggregate localization results (A' metric) for MVAB, sLORETA, dSPM, two variants of MCE (MCE_1 and MCE_2), and the three variants of Champagne ($CHAMP_S$, $CHAMP_D$, and $CHAMP_M$) for recovering three correlated sources with unknown orientations. Right: Estimated time-course correlation coefficient results for the eight algorithms. Error bars denote the standard error over 100 simulations.

three-dimensional nearest neighbors and is at least 10% of the maximum activation of the image. (This thresholding at 10% is designed to filter out any spurious peaks or ripples in the image that are much weaker than the maximum peak.)

After all the local peak locations were obtained, we tested whether each true source location has at least one local peak within a centimeter of it. (We also tested the performance at a radius of half a centimeter and the trends were similar for all algorithms.) If there were multiple local peaks within one centimeter of a particular true source, we counted it as one hit. Each image has the possibility of having 0, 1, 2, or 3 hits and we divided the number of hits by three to get the HR. Determining the false positive rate is more tenuous. There

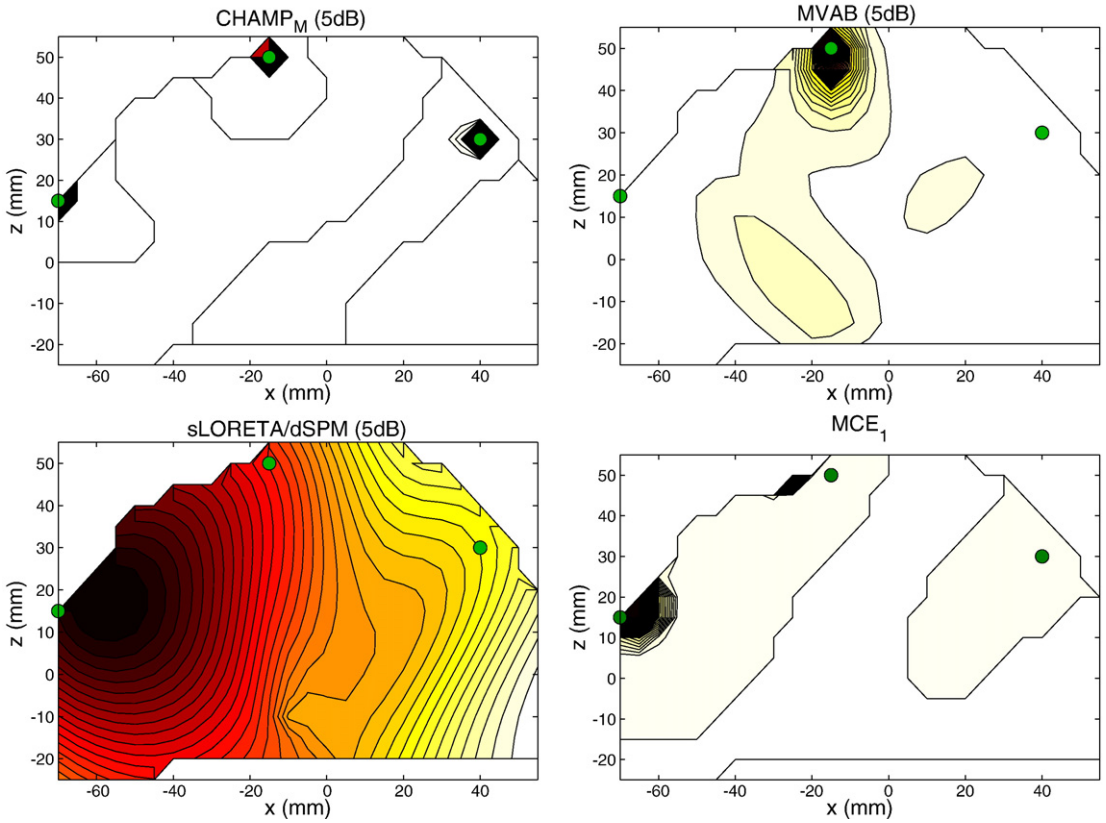

Fig. 2. Localization of three highly correlated dipoles at SNIR = 5 dB, the green circles denote the true locations of the sources and the red-to-white colored surface plot shows the maximum-intensity projection of the source power estimates produced by each algorithm. The color scale gradation goes from dark being the maximum to light being the minimum.

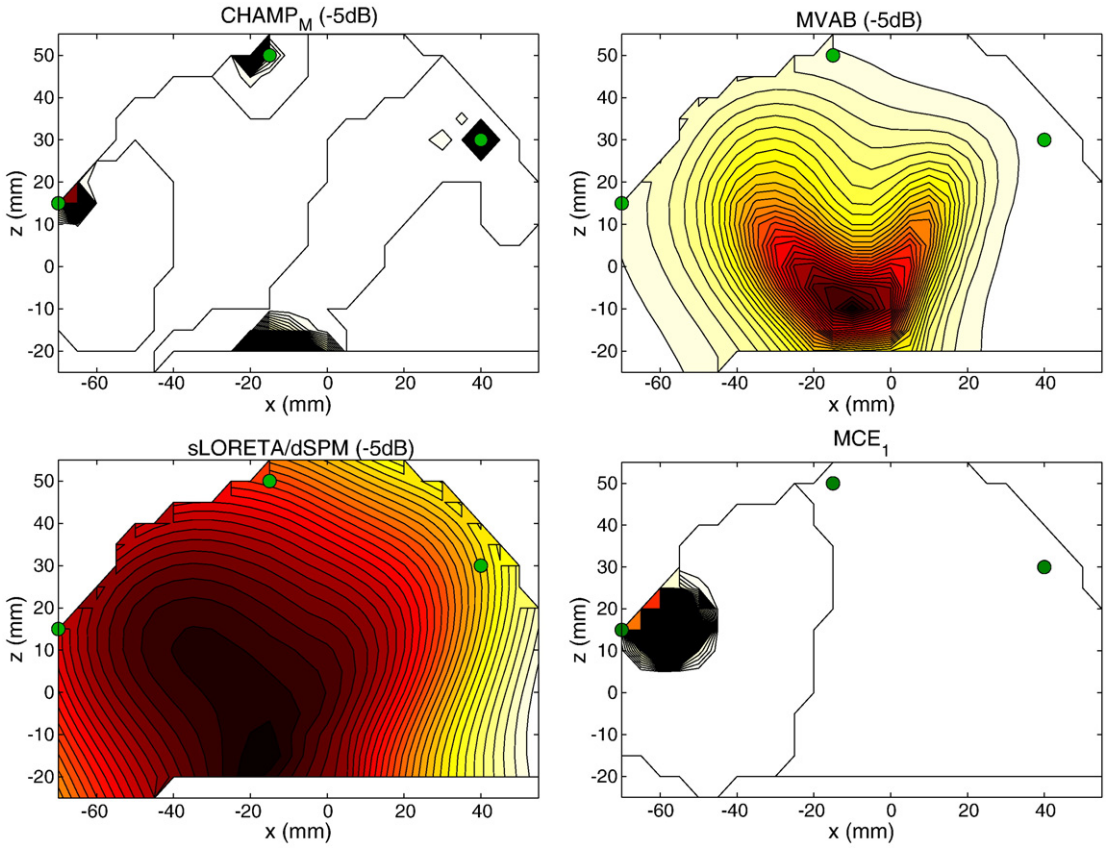

Fig. 3. Localization of 3 highly correlated dipoles at SNIR = −5 dB, the green circles denote the true locations of the sources and the red-to-white colored surface plot shows the maximum-intensity projection of the source power estimates produced by each algorithm. The color scale gradation goes from dark being the maximum to light being the minimum.

is not a clear maximum number of possible false positives, as there is with hits. We decided that given the spatial smoothness of the images, we could place a ceiling on the number of local peaks in an image at 5% of the total number of voxels or 100 voxels in the case of our simulations. Thus we divided the number of false positives, those local peaks not within one centimeter of a seeded source location, by 100 to get the FR. Since each algorithm was treated the same for the determination of the FR, we do not think that this estimated ceiling on the false positives caused any bias in our performance evaluation.

For evaluating the time-course reconstruction we used the correlation coefficient between the true time course seeded for the simulations and the estimated time series from each algorithm. Both the A' metric and correlation coefficient range from 0 to 1, with 1 implying perfect localization and time-course estimation. For both the localization accuracy and the time-course estimation assessments 100 simulations were averaged. Fig. 1 displays comparative results for the eight algorithms at different SNIR levels with standard errors. Fig. 1 left demonstrates the A' metric and Fig. 1 right shows the time-course

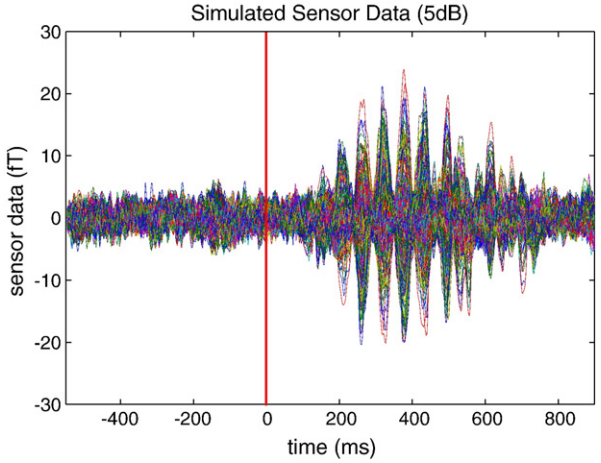

Fig. 4. Sensor data from all 275 MEG channels associated with Fig. 2 example. The red line denotes the pre- and post-stimulus periods.

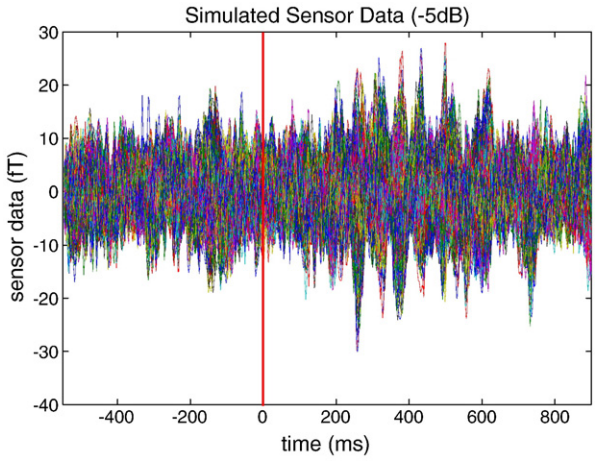

Fig. 5. Sensor data from all 275 MEG channels associated with Fig. 3 example. The red line denotes the pre- and post-stimulus periods.

correlation coefficients. All three variants of Champagne quite significantly outperform the others. For all further analyses, only $CHAMP_M$ was used because of its superior performance.

Figs. 2 and 3 show sample reconstructions of three correlated dipoles at 5 dB and −5 dB SNIR respectively. The associated sensor data is depicted in Figs. 4 and 5. Inter- and intra-dipole correlation

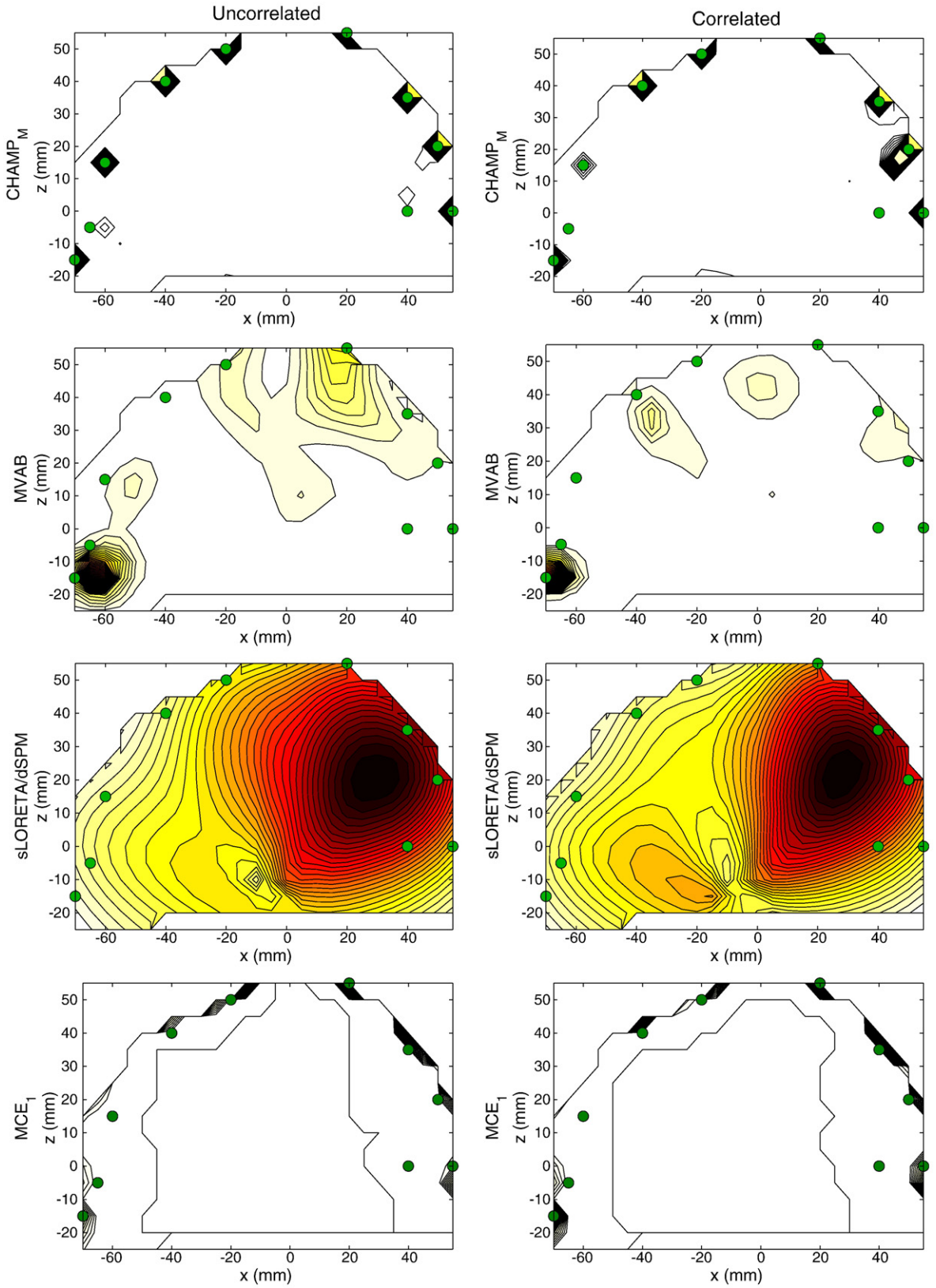

Fig. 6. Localization of 10 dipoles. Left column: Uncorrelated sources; right column: correlated sources. The green circles denote the true locations of the sources and the red-to-white colored surface plot shows the maximum-intensity projection of the source power estimates produced by each algorithm. The color scale gradation goes from dark being the maximum to light being the minimum.

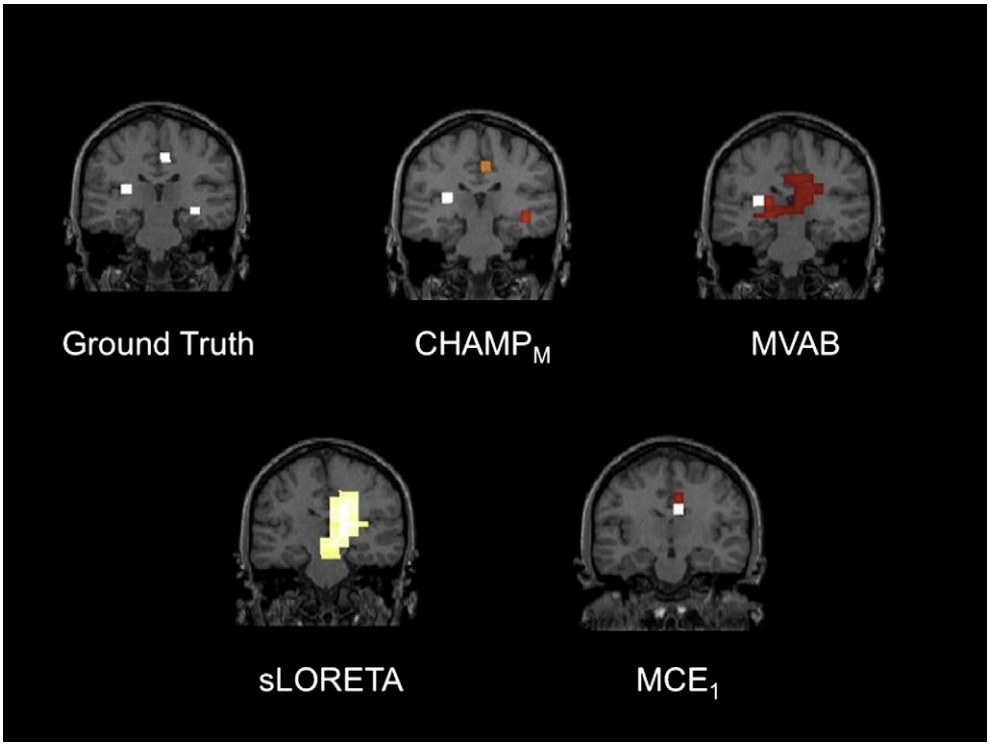

Fig. 7. Deep source example: 3 dipoles were seeded close to the center of the head to test the performance of the algorithm on deep sources. Champagne does the best at recovering these three sources.

coefficients were 0.90 in each case. All of the surface plots are maximum-intensity projections of estimated power maps onto the coronal plane. These power maps were obtained by calculating the average power at every voxel from the reconstructed voxel time courses. The green circles indicate the projection of the true locations of the sources and the surface plot shows the maximum-intensity projection of the power-map. For these images the black/dark red regions are the maximum and the minimum is pale yellow/white. The figures show reconstructions from CHAMP$_M$ against MVAB, MCE$_1$, and sLORETA/dSPM. In practice, the localization maps for sLORETA and dSPM are nearly identical (e.g., see Fig. 1), so we chose to only show the sLORETA images to represent both methods. CHAMP$_M$ performs significantly better at localizing the activity at both high and low SNIR levels with highly correlated dipoles.

Fig. 6 shows a sample reconstruction of a much more complex source configuration involving 10 dipolar sources. We compared CHAMP$_M$ versus MVAB, MCE$_1$, and sLORETA/dSPM assuming (i) the dipoles had zero inter- and intra-dipole correlations, and (ii) they had a correlation coefficient of 0.5 for both inter- and intra-dipole interactions. The SNIR for these simulations was 10 dB. CHAMP$_M$ performs significantly better than the other algorithms on both the correlated and uncorrelated cases. Note that the reason superficial sources are easier to find via Champagne (and with many other methods) is simply because the effective SNIR of deep sources is

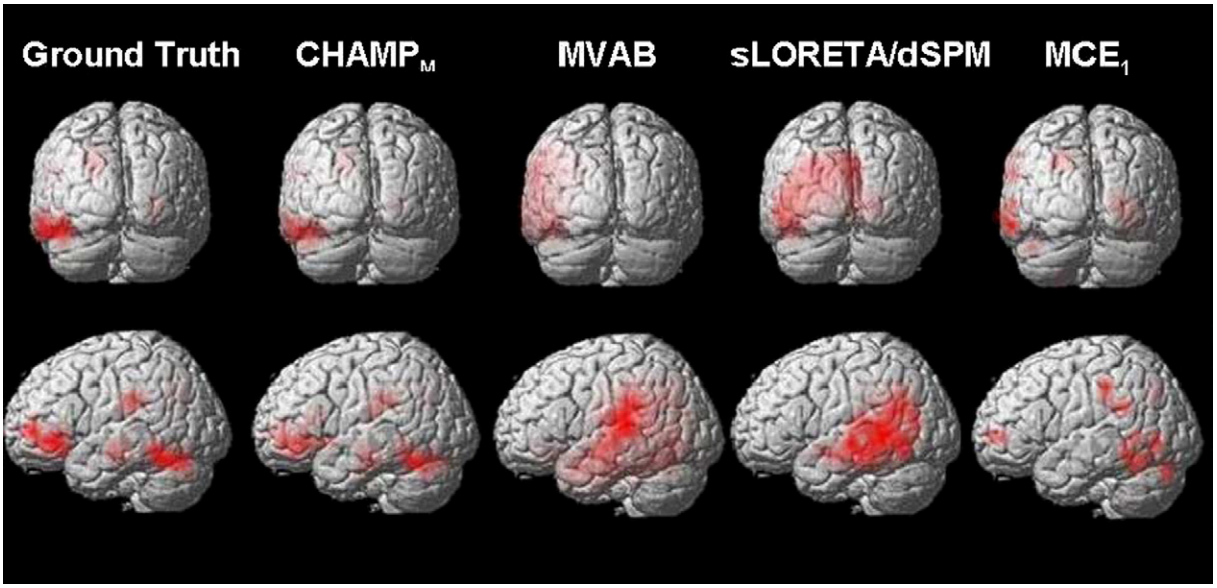

Fig. 8. Localization of diffuse sources. 10 clusters of 10 dipoles each were seeded in the brain and reconstructed using various algorithms. True and estimated sources were then projected to the rendered surface of the brain. Champagne comes the closest to matching the true sources.

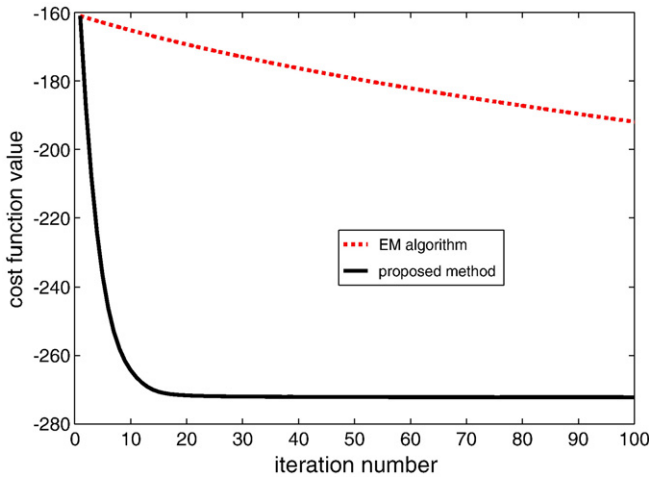

Fig. 9. Convergence rate of proposed update rules relative to a conventional EM implementation based on Friston et al. (2008); Sato et al. (2004) and Wipf (2006).

significantly lower. This occurs because deep sources, when combined with superficial ones, contribute very little to the numerator of Eq. (19) because the norm of the associated lead-field columns is relatively small. It is not because there is an intrinsic bias favoring superficial dipoles; in fact, the global minimum of $\mathcal{L}(\Gamma)$ is invariant to the lead-field column norms in the sense that the activation/sparsity pattern is completely unaffected.

As evidence that deeper sources need not necessarily be more difficult to find for a fixed SNIR, consider a scenario with three sources that are all reasonably deep. In this situation, the relative SNIR of each source will be comparable, meaning the overall source power is somewhat evenly distributed. If we are assuming a fixed SNIR for three superficial sources versus three deep ones, then the power associated with the deep ones will necessarily be larger to compensate for the small lead-field values. As a result, these deep sources will be of similar difficulty to find. So it is really in the situation where the SNIR is fixed but both deep and superficial sources are present that trouble arises. The shallow sources will dominate the SNIR calculation and will be readily located, while the deep ones contribute almost nothing to the data covariance and will likely be overlooked.

Fig. 7 shows recovery results using three deep sources with 0.5 correlation (inter- and intra-dipole) and 2 dB SNIR, with interference from real brain noise as in previous experiments. Clearly Champagne has no difficulty in this scenario while the other algorithms struggle. Of course in a practical situation, the SNIR for deep sources may be well below 2 dB, rendering them more difficult to find.

We also tested Champagne on more distributed source configurations. Previously, we created sources constrained to one voxel; now we consider sources formed from activity in 10 adjacent voxels to form clusters. We seeded 10 of these source clusters throughout the brain volume of interest. The lead-field used for this experiment is the same as the previous experiments with a spatial resolution of 5 mm. The SNIR was 10 dB and the intra-dipole correlation was 0.5 in all cases. The inter-dipole correlation was 0.8 for dipoles within the same cluster and 0.5 between dipoles in different clusters. We visualized the clusters by projecting the source power estimates to the surface of the 3-D rendered MRI.

From Fig. 8 we observe that Champagne was able to resolve much of each cluster and 80 of the 100 voxels were found to be significantly active above any small spurious peaks. Comparing with the ground truth plot, Champagne was able to reconstruct the clusters quite accurately, while MCE, sLORETA, and MVAB are not able to do so. Consequently, Champagne is found to be both applicable to focal sources as well as small diffuse clusters.

Finally, Fig. 9 gives an example of the improvement in convergence rate afforded by our method relative to an EM implementation analogous to (Friston et al., 2008; Sato et al., 2004); the per-iteration cost is the same for both methods. A simulated 2D dipolar source was generated and projected to the sensors using the experimental

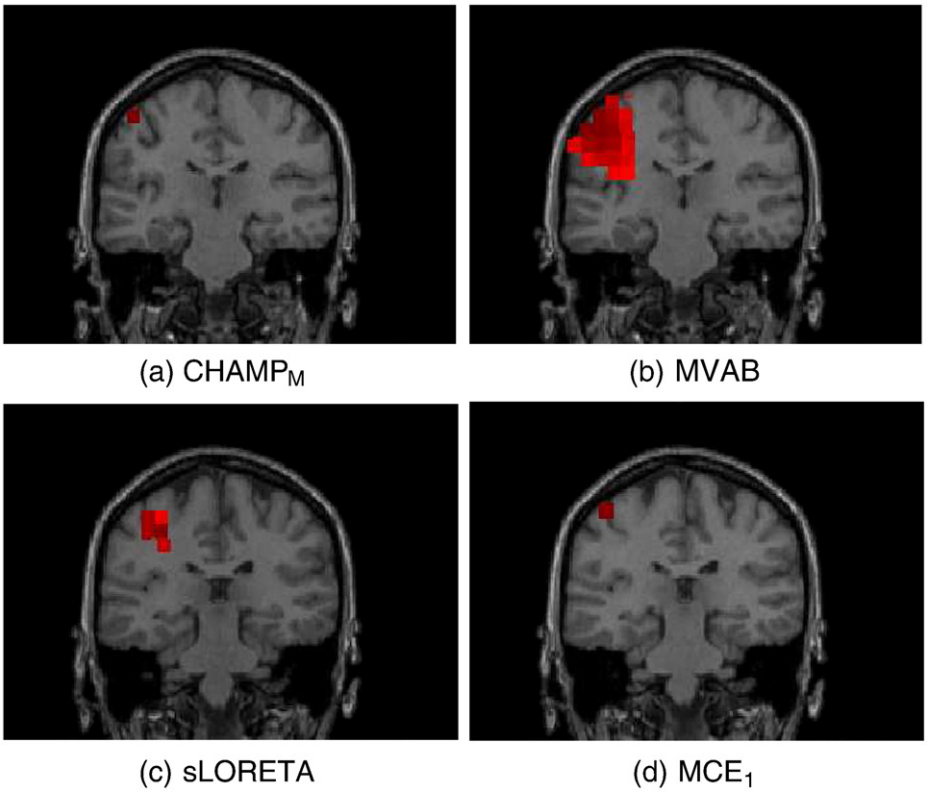

Fig. 10. Localization of SEF data. All four algorithms localize to somatosensory cortical areas, but Champagne and MCE are the most focal.

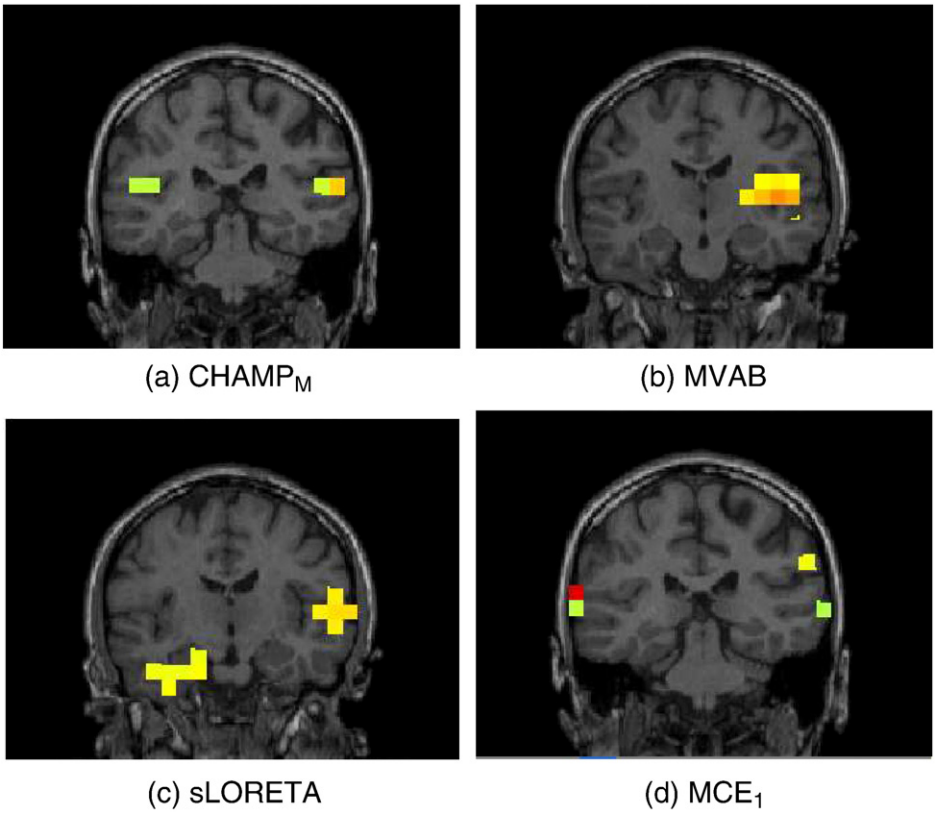

Fig. 11. Localization of AEF data (subject 1). Only Champagne is able to localize bilateral activity in primary auditory cortex.

paradigm described in Zumer et al. (2007) with $d_s = 1725$ voxels. The signal was corrupted by 10 dB additive Gaussian sensor noise. Fig. 9 displays the reduction in the Champagne cost function $\mathcal{L}(\Gamma)$ as a function of the iteration number. Consistent with previous observations, the EM updates are considerably slower in reducing the cost. While the detailed rationale for this performance discrepancy is

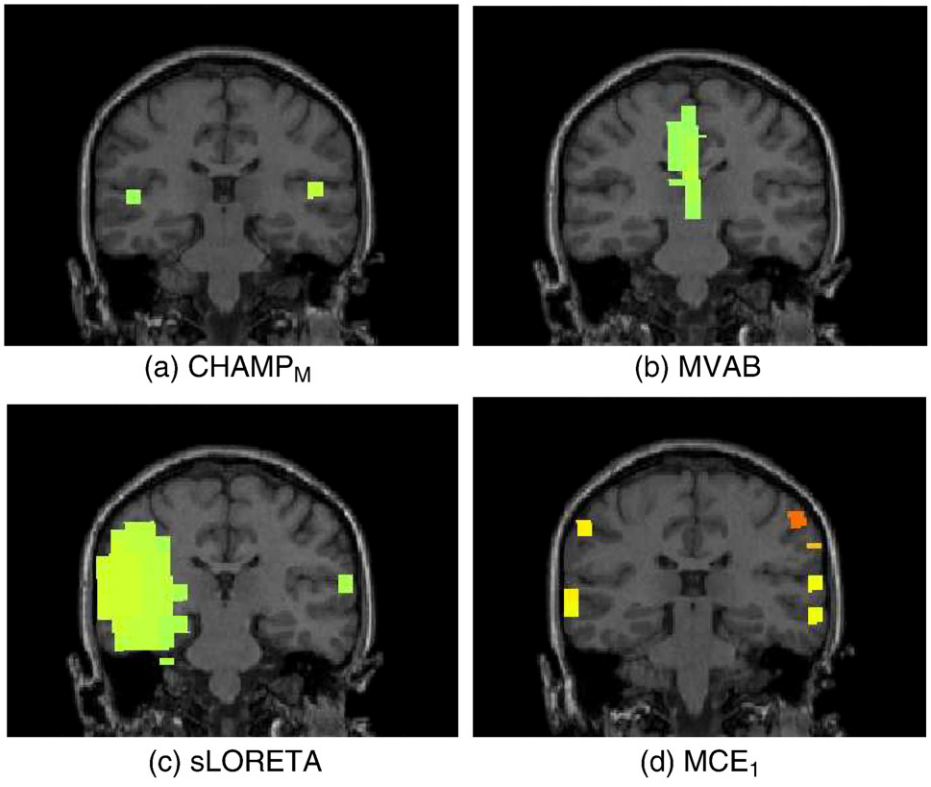

Fig. 12. Localization of AEF data (subject 2). Only Champagne is able to localize bilateral activity in primary auditory cortex.

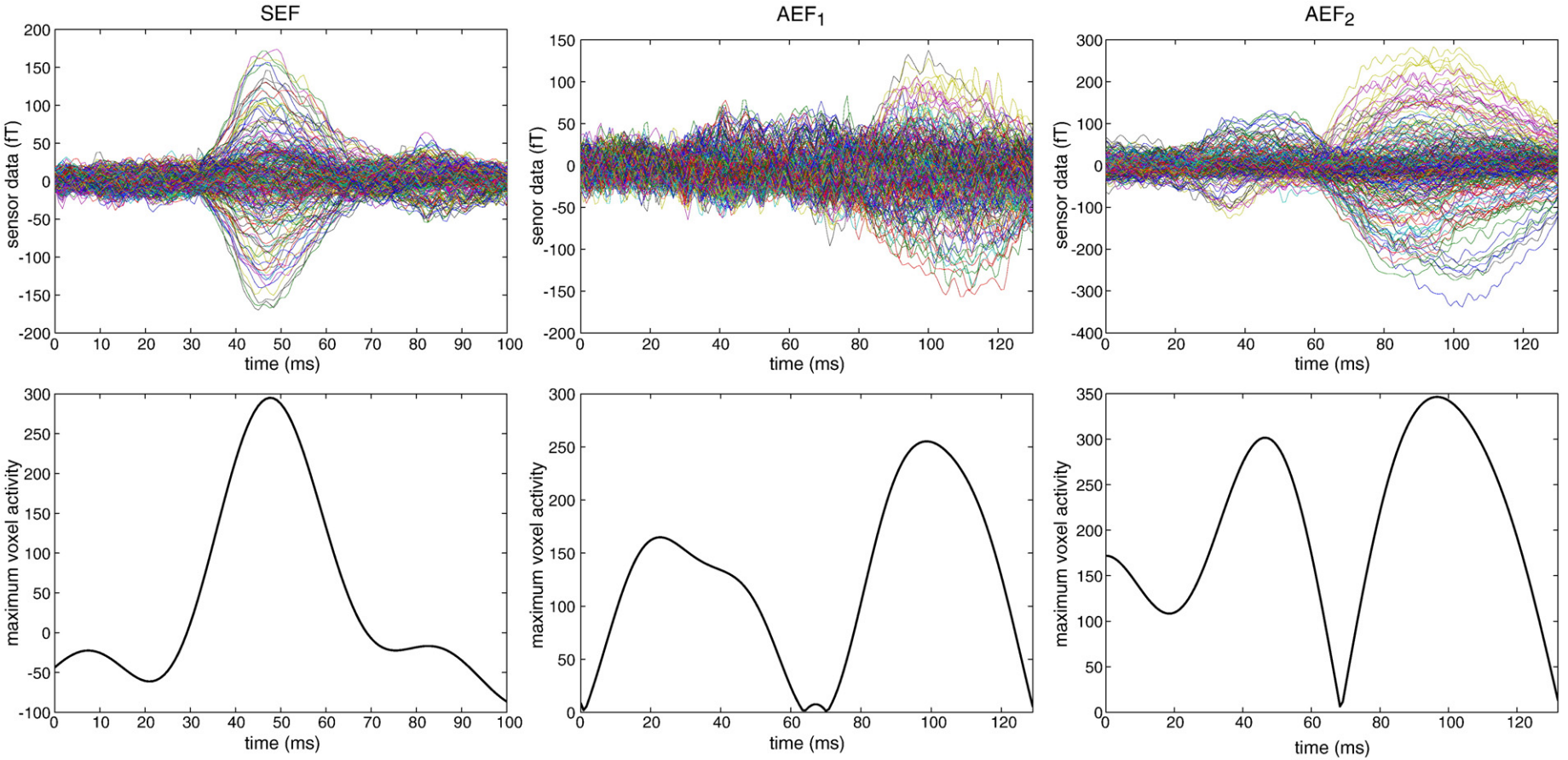

Fig. 13. Top row: Sensor data from all 275 MEG channels associated with Figs. 10, 11, and 12. Bottom row: Recovered time courses at the maximum-intensity voxels as estimated via CHAMP$_M$. Left: Left somatosensory cortex showing peak at 50 ms. Middle: Left auditory cortex from subject 1 (right auditory cortex, not shown, is similar) showing M50 and M100. Right: Right auditory cortex from subject 2 (left auditory cortex, not shown, is similar) showing M50 and M100.

beyond the scope of this paper, ultimately it is a consequence of the different underlying bounds being used to form auxiliary functions. EM leads to slower convergence because it is effectively using a much looser bound around zero than the bound described in the Learning the hyperparameters section and therefore fails to fully penalize redundant or superfluous components. This prevents the associated hyperparameters from going to zero very quickly, drastically slowing the convergence rate. More details on this topic can be found in Wipf and Nagarajan (2009).

Real data

Three stimulus evoked data sets were collected from normal, healthy research subjects on a 275-channel CTF System MEG device. The first data set was a sensory evoked field (SEF) paradigm, where the subject's right index finger was tapped for a total of 240 trials. A peak is typically seen 50 ms after stimulation in the contralateral (in this case, the left) somatosensory cortical area for the hand, i.e., dorsal region of the postcentral gyrus. Champagne and MCE_1 were able to localize this activation to the correct area of somatosensory cortex as seen in Fig. 10 and the estimated time course shows the typical 50 ms peak Fig. 13 (left). The other algorithms were also able to localize this activity, but the estimated activations are much more diffuse. Champagne and MCE's sparsity characteristics result in focal activations that do not have spatial blur like MVAB and sLORETA/dSPM. Note that for these examples, we do not perform a maximum-intensity projection. Instead, we simply find the voxels with maximum power (obtained from the estimated time courses). Champagne and MCE's activations were very focal, but for the other two algorithms we thresholded the image of estimated source power to obtain a focal activation around the maxima.

The other two data sets analyzed were from an auditory evoked field (AEF) paradigm, where two subjects were presented tones binaurally for a total of 120 trials. There are two typical peaks seen after the presentation of an auditory stimulus, one at 50ms and one at 100 ms, called the M50 and M100 respectively. The auditory processing of tones is bilateral at early auditory cortical areas and the activations are correlated. Champagne was able to localize activity in both primary auditory cortices for both subjects, Figs. 11 and 12, and the time courses for these two activations reveal the M50 and M100 Fig. 13 (middle) and (left). In general, these figures demonstrate that Champ_{M} outperforms both MVAB, MCE, and sLORETA/dSPM. The analysis of simple auditory paradigms is problematic for MVAB because it cannot handle the bilateral correlated sources well. sLORETA/dSPM does show some degree of bilateral activation but there is a large amount of spatial blur, while Champagne is focal in its localization. Note that the first iteration of Champagne is very related to sLORETA/dSPM (see Wipf et al., 2007) so this result is perhaps not surprising. Finally, MCE like Champagne produces focal estimates, but with additional spurious peaks and bilateral activations outside of auditory cortical areas.

Conclusion

This paper derives a novel empirical Bayesian algorithm for MEG source reconstruction that readily handles multiple correlated sources with unknown orientations, a situation that commonly arises even with simple imaging tasks. Based on a principled cost function and fast, convergent update rules, this procedure displays significant theoretical and empirical advantages over many existing methods. We have restricted most of our exposition and analyses to MEG; however, preliminary work with EEG is also promising. For example, on a real-world passive visual task where subjects viewed flashing foreground/background textured images, our method correctly localizes activity to the lateral occipital cortex while two state-of-the-art beamformers fail. This remains an active area of research.

There are a number of directions for future research. First, while not the focus of this paper, our model can readily be augmented to produce model evidence approximations, which have been proposed for Bayesian model comparison/selection purposes (Friston et al., 2008, 2006; Wipf and Nagarajan, 2009). While here such approximations emerge from the concavity bounds associated with the derivation of Champagne in the Learning the hyperparameters section, in Friston et al. (2008, 2006), related bounds follow from factorial (or mean-field) assumptions built into a variational free energy framework. The connection between these two different types of bounds, as well as their relative performance in model selection tasks, has not been explored to our knowledge.

Secondly, a full investigation of the effects of the temporal windowing of sensor time courses, and algorithmic extensions for learning optimal windows, is very important to empirical Bayesian methods such as Champagne. There is an intrinsic trade-off here. Large windows are most effective for localizing stationary sources that remain active across the whole window length; however, if a source has limited temporal extent or is moving, extending the window size can drastically reduce performance. In contrast, small windows are optimal for non-stationary, ephemeral sources, but the effective SNIR will generally be worse. In preliminary studies with data from facial processing experiments, adjusting the window size has been crucial for localizing regions such as the fusiform face area. This issue is implicitly considered in Bolstad et al. (2009), where temporal basis functions are used in extended MCE framework (a notion that could potentially be applied to Champagne as well). Spatial basis functions can also be incorporated to allow more flexible estimation of distributed sources, a useful notion that has been applied in a wide variety of settings (Bolstad et al., 2009; Limpiti et al., 2006; Phillips et al., 2002; Ramírez and Makeig, 2006; Wipf and Nagarajan, 2009).

Acknowledgments

This work was funded by NIH grants R01 DC006435 and DC004855 to SSN.

Appendix A. Introduction to conjugate duality

At the heart of the algorithm derived in the Algorithm derivation section is the ability to represent a concave function in its dual form. For example, given a concave function $f(y) : \mathbb{R} \rightarrow \mathbb{R}$, the dual form is given by

$$f(y) = \inf_{\upsilon} \left[\upsilon y - f^*(\upsilon)\right], \tag{20}$$

where $f^*(\upsilon)$ denotes what is called the *conjugate function* (Boyd and Vandenberghe, 2004). Geometrically, this can be interpreted as representing $f(y)$ as the lower envelope or infimum of a set of lines parameterized by υ. The selection of $f^*(\upsilon)$ as the intercept term ensures that each line is tangent to $f(y)$. If we drop the maximization in Eq. (20), we obtain the bound

$$f(y) \leq \upsilon y - f^*(\upsilon). \tag{21}$$

Thus, for any given υ, we have a rigorous bound on $f(y)$; we may then optimize over υ to find the optimal or tightest bound in a region of interest. In higher dimensions, the strategy is the same, only now the bounds generalize from lines to hyperplanes.

Appendix B. Proof of Theorem 1

We begin with Property 1. While it is possible to prove this result using manipulations of $\mathcal{L}(\Gamma)$ in Γ-space, it is convenient to transform

the problem to an equivalent one in S-space. Specifically, the solution S^* will be the global minimum of the dual optimization problem

$$S^* = \lim_{\Sigma_\varepsilon \to 0} \left[\min_{S: B = LS} g(S) \right] \tag{22}$$

where

$$g(S) \triangleq \min_{\Gamma} \left[\sum_{i=1}^{d_s} \text{trace}\left(S_i^T \Gamma_i^{-1} S_i \right) + \log|\Sigma_b| \right] \tag{23}$$

and $\Sigma_b = \Sigma_\varepsilon + L\Gamma L^T$ as defined in the main text. The global and local minima of Eq. (22) correspond with those of $\mathcal{L}(\Gamma)$ by straightforward extension of results in Wipf and Nagarajan (2008). Therefore, we only need to show that the global minimizer of Eq. (22) S^* satisfies Property 1. For simplicity, we will assume that $\sum_\varepsilon = \varepsilon I$ with $\varepsilon \to 0$ and that spark $(L) = d_b + 1$ (i.e., every subset of d_b lead-field columns are linearly independent). The more general case is straightforward to handle but complicates notation and exposition unnecessarily.

To begin we require two intermediate results for convenience and to avoid confusion, we define S_{gen} as the matrix of generating sources that we would like to recover.

Lemma 1. *The function $g(S)$ satisfies*

$$g(S) = O(1) + (d_b - \min[d_b, Dd_c]) \log \varepsilon, \tag{24}$$

where D equals the cardinality of the set $\{S_i : ||S_i|| > 0\}$, i.e., D is the number of S_i not equal to zero.[2,3]
In words this result states that $g(S)$ will be $O(1)$ unless $Dd_c < d_b$, in which case $g(S)$ will be dominated by the log ε term when ε is small.

Proof. Computing $g(S)$ via Eq. (23) involves a minimization over two terms. The first term encourages each Γ_i to be large, the second encourages each Γ_i to be small. Whenever a given $S_i = 0$, the first term can be ignored and the associated Γ_i is driven to exactly zero by the second term. In contrast, for any $S_i \neq 0$, the minimizing Γ_i can never be zero for any $\varepsilon \geq 0$ or the first term will be driven to infinity. This a manifestation of the fact that

$$\arg\min_{x \geq 0} \left[\frac{1}{x} + \log(x + \varepsilon) \right] > 0, \quad \forall \varepsilon \geq 0. \tag{25}$$

Consequently, for any given S, the associated minimizing Γ will necessarily have a matching sparsity profile, meaning the indices of zero-valued S_i will align with zero-valued block-diagonal elements in Γ.[4]

Whenever $Dd_c < d_b$, the above analysis (and the assumption that spark $(L) = d_b + 1$) ensures that the minimizing Σ_b will be full rank even for $\varepsilon = 0$. This implies that $g(S) = O(1)$ for essentially the same reason that

$$\min_{x \geq 0} \left[\frac{1}{x} + \log(x + \varepsilon) \right] = O(1). \tag{26}$$

In contrast, when $Dd_c < d_b$, the minimizing Σ_b will become degenerate when $\varepsilon \to 0$. Let λ_i denote the i-th non-zero eigenvalue of $L\Gamma L^T$ at the minimizing Γ. The spark assumption (coupled with the analysis above) guarantees that there will be Dd_c such eigenvalues. Then we have

$$\log|\Sigma_b| = \sum_{i=1}^{Dd_c} \log(\lambda_i + \varepsilon) + (d_b - Dd_c) \log \varepsilon. \tag{27}$$

This gives $g(S) = O(1) + (d_b - Dd_c) \log \varepsilon$. ■

Lemma 2. *For any solution S such that $B = LS$, D will always satisfy $D \geq d_a$. The sources S that achieve equality are unique and satisfy $S = S_{gen}$.*

Proof. This result represents a simple extension of uncertainty principles detailed in Cotter et al. (2005) and Donoho and Elad (2003). In particular, based on Lemma 1 in Cotter et al. (2005), if S_{gen} satisfies

$$d_a d_c < (\text{spark}(L) + \text{rank}(B) - 1)/2 \leq (\text{spark}(L) + \min[d_t, d_a d_c] - 1)/2, \tag{28}$$

then no other solution S can exist such that $B = LS$ and $D \geq d_a$. Additionally, by directly applying results from Elad (2006), we find that this condition will also hold for all S, except a set with measure zero, if

$$d_a d_c < \text{spark}(L) - 1. \tag{29}$$

■

Given these two results, the proof of Property 1 is simple. In the limit as $\varepsilon \to 0$, Lemma 1 dictates that $g(S)$ is minimized when D is minimized. Lemma 2 then shows that D is uniquely minimized (with $D \geq d_a$) when $S = S_{gen}$.

Property 2 is relatively easy to show by leveraging Theorem 2 from Wipf and Nagarajan (2007). For our purposes, this result implies that if the data covariance C_b satisfies $C_b = \Sigma_b$ for some Γ, then the cost function $\mathcal{L}(\Gamma)$ is unimodal and $C_b = \Sigma_b$ at any minimizing solution. When $\varepsilon \to 0$, the data covariance satisfies

$$C_b = \frac{1}{d_t} BB^T = \frac{1}{d_t} LS_{gen} S_{gen}^T L^T = L\left(\frac{1}{d_t} S_{gen} S_{gen}^T \right) L^T. \tag{30}$$

Given the conditions of Property 2, $S_{gen} S_{gen}^T$ will be zero except for $d_c \times d_c$ block-diagonal elements. Therefore, if $\Gamma = 1/d_t S_{gen} S_{gen}^T$ (which is allowable given the specified block-diagonal structure of Γ), then $\Sigma_b = L\Gamma L^T = C_b$ and we have no local minima.

References

Baillet, S., Mosher, J., Leahy, R., Nov. 2001. Electromagnetic brain mapping. IEEE Signal Process. Mag. 14–30.
Berger, J.O., 1985. Statistical Decision Theory and Bayesian Analysis, 2nd edition. Springer-Verlag, New York.
Bolstad, A., Van Veen, B., Nowak, R., 2009. Space-time event sparse penalization for magneto-/electroencephalography. NeuroImage 46 (4), 1066–1081.
Boyd, S., Vandenberghe, L., 2004. Convex Optimization. Cambridge University Press.
Cotter, S.F., Rao, B.D., Engan, K., Kreutz-Delgado, K., April 2005. Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans. Signal Process. 53 (5), 2477–2488.
Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., April 2000. Dynamic statistical parametric mapping: combining fMRI and MEG for high-resolution imaging of cortical activity. Neuron 26 (1), 55–67.
Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., Leahy, R.M., Sept. 2004. Mapping human brain function with MEG and EEG: methods and validation. NeuroImage 23 (Suppl. 1), S289–S299.
Donoho, D.L., Elad, M., March 2003. Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization. Proc. Natl. Acad. Sci. 100 (5), 2197–2202.
Elad, M., 2006. Sparse representations are most likely to be the sparsest possible. EUROSIP J. Appl. Signal Process. 1–12.
Friston, K., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the MEG/EEG inverse problem. NeuroImage 39 (1), 1104–1120.
Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2006. Variationl free energy and the Laplace approximation. NeuroImage 34, 220–234.

[2] Here we have adopted the notation $f(x) = O(h(x))$ to indicate that $|f(x)| < C_1|h(x)|$ for all $x < C_2$, with C_1 and C_2 constants independent of x.
[3] If $S = S_{gen}$, then by definition $D = d_a$.
[4] This point can be made more rigorous as shown in the first author's PhD thesis, but we omit lengthy details here.

Gorodnitsky, I., George, J., Rao, B., Oct. 1995. Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. J. Electroencephalogr. Clin. Neurophysiol. 95 (4), 231–251.

Huang, M., Dale, A., Song, T., Halgren, E., Harrington, D., Podgorny, I., Canive, J., Lewis, S., Lee, R., 2006. Vector-based spatial–temporal minimum ℓ_1-norm solution for MEG. NeuroImage 31 (3), 1025–1037.

Limpiti, T., Veen, B.V., Wakai, R., Sept. 2006. Cortical patch basis model for spatially extended neural activity. IEEE Trans. Biomed. Eng. 53 (9), 1740–1754.

Mattout, J., Phillips, C., Penny, W., Rugg, M., Friston, K., 2006. MEG source localization under multiple constraints: an extended Bayesian framework. NeuroImage 30, 753–767.

Nagarajan, S.S., Attias, H.T., Hild, K.E., Sekihara, K., Sept. 2007. A probabilistic algorithm for robust interference suppression in bioelectromagnetic sensor data. Stat. Med. 26 (21), 3886–3910.

Nummenmaa, A., Auranen, T., Hämäläinen, M., Jääskeläinen, I., Lampinen, J., Sams, M., Vehtari, A., 2007. Hierarchical Bayesian estimates of distributed MEG sources: theoretical aspects and comparison of variational and MCMC methods. NeuroImage 35 (2), 669–685.

Pascual-Marqui, R.D., 2002. Standardized low resolution brain electromagnetic tomography (sloreta): technical details. Methods Find. Exp. Clin. Pharmacol. 24 (Suppl. D), 5–12.

Phillips, C., Mattout, J., Rugg, M., Maquet, P., Friston, K., January 2005. An empirical Bayesian solution to the source reconstruction problem in EEG. NeuroImage 24, 997–1011.

Phillips, C., Rugg, M., Friston, K., July 2002. Anatomically informed basis functions for EEG source localization: combining functional and anatomical constraints. NeuroImage 13 (3), 678–695 pt. 1.

Ramírez R. and Makeig S., June 2006. "Neuroelectromagnetic source imaging using multiscale geodesic neural bases and sparse Bayesian learning," Human Brain Mapping, 12th Annual Meeting, Florence, Italy.

Sahani, M., Nagarajan, S.S., 2004. Reconstructing MEG sources with unknown correlations. Adv. Neural Inf. Process. Syst. 16.

Sarvas, J., 1987. Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem. Phys. Med. Biol. 32, 11–22.

Sato, M., Yoshioka, T., Kajihara, S., Toyama, K., Goda, N., Doya, K., Kawato, M., 2004. Hierarchical Bayesian estimation for MEG inverse problem. NeuroImage 23, 806–826.

Sekihara, K., Nagarajan, S.S., 2008. Adaptive Spatial Filters for Electromagnetic Brain Imaging. Springer.

Sekihara, K., Sahani, M., Nagarajan, S.S., 2005. Localization bias and spatial resolution of adaptive and non-adaptive spatial filters for MEG source reconstruction. NeuroImage 25, 1056–1067.

Snodgrass, J.G., Corwin, J., 1988. Pragmatics of measuring recognition memory: applications to dementia and amnesia. J. Exp. Psychol.: Gen. 17 (1), 34–50.

Uutela, K., Hamalainen, M., Somersalo, E., 1999. Visualization of magnetoencephalographic data using minimum current estimates. NeuroImage 10, 173–180.

Wipf D.P., 2006. "Bayesian methods for finding sparse representations," PhD Thesis, University of California, San Diego.

Wipf D.P. and Nagarajan S., June 2007. "Beamforming using the relevance vector machine," International Conference on Machine Learning.

Wipf, D.P., Nagarajan, S., 2008. A new view of automatic relevance determination. Advances in Neural Information Processing Systems 20. MIT Press.

Wipf, D.P., Nagarajan, S., February 2009. A Unified Bayesian Framework for MEG/EEG Source Imaging. NeuroImage 44 (3).

Wipf D.P. and Nagarajan S., 2009. "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," Submitted.

Wipf, D.P., Ramírez, R.R., Palmer, J.A., Makeig, S., Rao, B.D., 2007. Analysis of empirical Bayesian methods for neuroelectromagnetic source localization. Adv. Neural Inf. Process. Syst. 19.

Zumer, J.M., Attias, H.T., Sekihara, K., Nagarajan, S.S., 2007. A probabilistic algorithm integrating source locations and noise suppression for MEG and EEG data. Neuroimage 37, 102–115.