Realtime MEG source localization with realistic noise

Sung Chan Jun **Barak A. Pearlmutter** **Guido Nolte**
Department of Computer Science
University of New Mexico
Albuquerque, NM 87131
{junsc,bap,nolte}@cs.unm.edu

Abstract

Iterative gradient methods like Levenberg-Marquardt (LM) are in widespread use for source localization from electroencephalographic (EEG) and magnetoencephalographic (MEG) signals. Unfortunately LM depends sensitively on the initial guess, particularly (and counter-intuitively) at higher signal-to-noise ratios, necessitating repeated runs. This, combined with LM's high per-step cost, makes its computational burden quite high. To reduce this burden, we trained a multilayer perceptron (MLP) as a real-time localizer. We used an analytical model of quasistatic electromagnetic propagation through the head to map randomly chosen dipoles to sensor activities, and trained an MLP to invert this mapping in the presence of various sorts of noise. *With realistic noise, our MLP is about five hundred times faster than n-start-LM with $n = 4$ to match accuracies, while our hybrid MLP-start-LM is about four times more accurate and thirteen times faster than 4-start-LM.*

1 Introduction

Source localization of EEG and MEG signals is important in medical diagnosis of conditions like epilepsy, in surgical planning, and in neuroscience research. Assuming dipolar sources, there are a number of localization methods in use (Hämäläinen et al., 1993). Among them, optimization using an iterative gradient method like LM (Levenberg, 1944; Marquardt, 1963) is one of the best, in terms of accuracy and computational burden. However, gradient methods require both a differentiable forward model and an initial guess. As we shall see, the efficiency and accuracy of the most popular gradient method for this problem, LM, depends sensitively on the initial guess, particularly at higher S/N ratios.

There is therefore motivation to build faster and more accurate source localizers. This is particularly important for our real time MEG brain-computer interface system, as we need to localize BSS-separated components in real time.

Since it is easy to create synthetic data consisting of pairs of corresponding dipole locations and sensor signals, it is tempting to train a universal approximator to solve the inverse problem directly, *i.e.* to map sensor signals directly to the dipole location and moment. The multilayer perceptron (MLP) of Rumelhart et al. (1986) has been popular for this purpose.

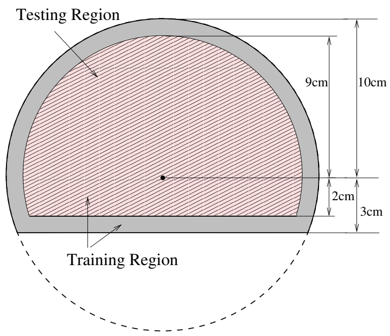
MLPs were first used for EEG dipole source localization and presented as feasible source localizers by Abeyratne et al. (1991), and a MLP structure composed of six separate networks was later used for EEG dipole localization by Zhang et al. (1998). Kinouchi et al. (1996) first used MLPs for MEG source localization by training on a noise-free dataset in spherical shell, while Yuasa et al. (1998) studied the two-dipole case for EEG dipole source localization under the assumption that each source dipole is in a restricted region. Hoey et al. (2000) investigated EEG measurements for both spherical and realistic head models, trained on a randomly generated noise-free dataset, and presented a comparison between MLP and iterative methods for localization with noisy signals at three fixed dipole locations. Sun and Sclabassi (2000) adapted an MLP to calculate forward EEG solutions from a spheroidal head model.

Here we train an MLP to localize dipoles from MEG measurements, and measure the efficacy of the resulting network under a variety of conditions. Its most interesting use is as a generator of the initial parameter values for LM, a role in which it excels.

2 Synthetic data

The synthetic data used in our experiments consists of corresponding pairs of dipole locations and moments, and sensor activations. The sensor activations are calculated by adding the results of a forward model and a noise model.

Dipoles within the training and testing set were drawn uniformly from a spherical region with a slice removed, as shown on the right. The training set used dipoles from the larger region, while the test set contained only dipoles from the smaller inner region.



2.1 Forward model

We use a standard analytic forward model of quasistatic electromagnetic propagation in a spherical head model (Sarvas, 1987; Mosher et al., 1999), with the sensor geometry of a 4D Neuroimaging Neuromag-122 gradiometer.

2.2 Noise model

For single trial data, the sensors in MEG systems have poor S/N ratios since MEG data is strongly contaminated not only by intrinsic sensor noise, but also by external fields, fields generated by various parts of the body (heart, eye muscles, retina), and parts of the brain not under study. Blind source separation of MEG data can drastically improve the situation by segregating noise from signal (Vigário et al., 1998; Tang et al., 1999), and the sensor attenuation vectors of the BSS-separated components can be well localized to equivalent current dipoles (Tang et al., 2000). However, the recovered field maps can be quite noisy, and conventional localization techniques require manual interaction.

In order to properly compare the performance of various localizers, we need a dataset for which we know the ground truth, but which contains the sorts of noise encountered in actual MEG recordings (Kwon et al., 2000). To this end, we created three noise processes with which to additively contaminate synthetic sensor readings. These are: uncorrelated Gaussian noise, correlated noise, and actual noise. The uncorrelated Gaussian noise is generated by simply drawing a Gaussian-distributed random number for each sensor. Correlated noise is made using the method of Lütkenhöner (1994),

1. equally well distribute 900 dipoles on a spherical surface, with dipole moments drawn from a zero-mean spherical Gaussian.
2. calculate a sensor activation through the analytic forward model for each dipole for each sensor and sum over all dipoles at each sensor.
3. scale the resultant sensor activation vector to yield a suitable RMS power.
4. use this vector as the noise.

Actual noise was taken from MEG recordings during periods in which the brain region of interest in the experiment was quiescent. These signals were not averaged. The actual noise, without scaling, has an RMS of roughly $P^n = 50$–100 fT/cm.

We scaled the additive noise to make the RMS power of the various sorts of noise equal. We measured the S/N ratio of a data set using the ratios of the powers in the signal and the noise: S/N (in dB) $= 10 \log_{10} P^s / P^n$ where P^s is the RMS (square root of mean square) of the sensor readings from the dipole and P^n is the RMS of the sensor readings from the noise.

3 Multilayer Perceptron

The MLP charged with approximating the inverse mapping had an input layer of 122 units, one for each sensor; two hidden layers with n_1 and n_2 units respectively, and an output layer of 6 units, representing the dipole location (x, y, z) and moment (m_x, m_y, m_z). The output units had a linear activation functions, while the hidden units had hyperbolic tangent activation functions to accelerate training (LeCun et al., 1991). All units had bias inputs, adjacent layers were fully connected, and there were no cut-through connections.

The 122 MEG sensor activations were scaled so that the RMS value was 0.5, and the corresponding dipole moment was scaled by the same factor. Dipole location and dipole moment parameters were further scaled down to ensure that they were under 80% of saturation of the output units.

The network weights were initialized with uniformly distributed random values between ± 0.1. Vanilla online backpropagation was used for training. No momentum was used, and η was chosen empirically.

3.1 MLP structural optimization

Beginning with intuitions drawn from the explorations of suitable numbers of hidden units by Hoey et al. (2000) for EEG localization, we empirically measured the tradeoff between approximation accuracy and computation time. Generalization was not a serious consideration, since training sets of arbitrary size could be easily constructed: our training sets ranged from 5,000–20,000, as circumstances dictated.

For practical reasons, we constrained our experiments to networks with no more than 110 hidden units in either hidden layer. We varied the number of hidden units in each layer from 10 to 110, in steps of 10, with $n_1 \geq n_2$. Each MLP was trained with a noise-free training dataset of 5,000 training exemplars, and the mean localization error for a noise-free test dataset of 2,500 was measured after 500 epochs of training. Training each network took up to two hours on an 800 MHz AMD Athlon computer. For each size, five runs were performed and the errors averaged.

The calculation time was measured in terms of equivalent adds for a forward pass, *i.e.* a localization. The number of equivalent additions per addition, multiplication, and hyperbolic tangent were about 1, 3, and 33, as measured on the above CPU. The equivalent floating points additions for the 122–n_1–n_2–6 MLP structure is therefore $(122n_1 + n_1n_2 + 6n_2) + 3(n_1n_2 + 123n_1 + 7n_2 + 6) + 33(n_1 + n_2)$.

Figure 1 shows both average localization error for training and testing versus calculation time for a localization. Each point in the figure describes a different network architecture. When the number of additions is small localization error is high. Increasing the computation reduces the localization error. The accuracy levels off after a while, probably due to incomplete convergence of the network training. From this result we choose the suitable MLP size as 122–60–20–6, as indicted in Figure 1. For our purposes this seemed a reasonable tradeoff between computation and accuracy.

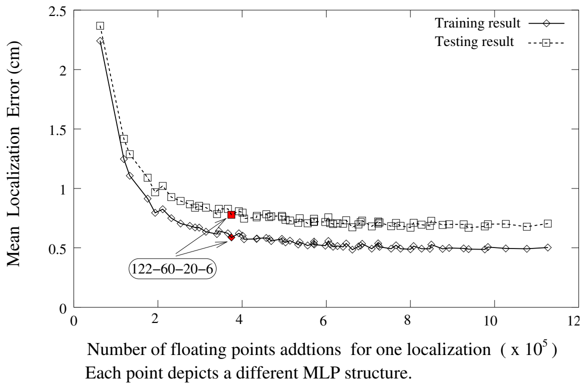


Figure 1: Mean localization error versus calculation time, as a parametric function of MLP structure. A noise-free synthetic dataset was used for both training and testing.

We constructed four training datasets, each with 20,000 exemplars, differing only in the type of noise: none, gaussian white, correlated, and actual (Section 2.2). No matter what a network was trained with, we always tested using actual noise. As described in Section 2, the S/N ratio was controlled by scaling the additive noise.

It took 500 epochs and about four hours on an 800 MHz AMD Athlon, to train a network of the selected architecture on a noisy dataset.

4 Effect of the initial guess on LM Localization

To see how the initial guess effects the LM localizer, we measured the localization performance of LM as a function of the distance from the initial dipole location to the actual location. The initial guess was chosen randomly on a sphere of radius d centered on the target.

For each S/N ratio, 300 noisy exemplars were created. The S/N ranged from 0–11 dB, and the distance d from 0–6 cm in steps of 1 cm. For each sensor activation and initial guess, LM finds the dipole parameters that minimize a quadratic function (Hämäläinen et al., 1993) of the difference between the predicted and input sensor activations,

$$c(\mathbf{x}) = ||\mathbf{B}(\mathbf{x}, \mathbf{Q}(\mathbf{x})) - \mathbf{B}_m||^2_{\Sigma^{-1}} \quad (1)$$

where $||\mathbf{v}||^2_{\mathbf{A}} = \mathbf{v}^T\mathbf{A}\mathbf{v}$, we define $\mathbf{B}(\mathbf{x}, \mathbf{Q}) = \mathbf{F}(\mathbf{x})\mathbf{Q}$, and $\mathbf{x}$ and $\mathbf{Q}$ denote a source dipole location vector and a source dipole moment vector, and $\mathbf{Q}$ can be expressed by the least square method as: $\mathbf{Q}(\mathbf{x}) = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{B}_m$.

$\mathbf{B}_m$ and $\mathbf{B}(\mathbf{x}, \mathbf{Q}(\mathbf{x}))$ are 122-element vectors with measured and calculated sensor activations through the forward model, respectively, and $\mathbf{F}(\mathbf{x})$ is the kernel of a spherical head model (Mosher et al., 1999). Σ is the noise covariance matrix, which is an identity matrix for spherical zero mean unit variance gaussian noise. If the noise is known, the covariance matrix can be easily calculated. However noise is generally unknown, so in reality people often assume a spherical covariance matrix. Alternatively, one can measure the sensor activations before stimulation or long after stimulation, and calculate the covariance matrix of those measurements. Since both of these techniques are popular, we simulate each. Fig-

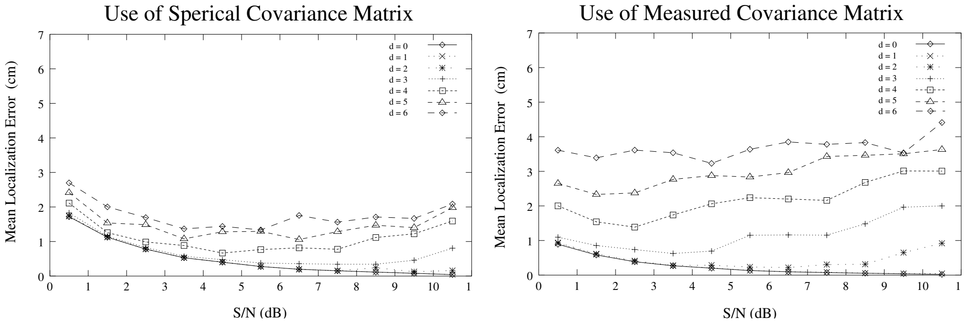


Figure 2: Mean localization error of 1-start-LM as a function of S/N at varying distances d between the initial guess and the actual source dipole. Left: spherical covariance matrix. Right: measured covariance matrix.

ure 2 shows the mean localization error for 300 test sets of varying S/N, for both spherical and empirical covariance matrices.

Figure 3 shows the mean localization error for 3,300 activations with various S/N as the distance between the target and the initial guess is varied. Figure 2 shows that the closer the initial guess, the better the performance, for the both covariance assumptions. With a good initial guess the empirical covariance yields much better localization performance, but at the expense of performance when the initial guess is further from the target.

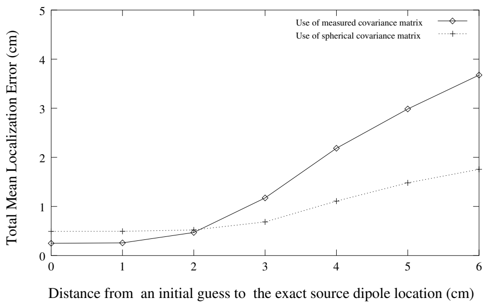


Figure 3: Localization error versus distance d between LM using spherical covariance and LM using measured covariance. S/N ranges from 0–11 dB.

These experiments show that signals having high S/N show a greater degradation as the initial guess is moved away from the target! One can attempt to compensate for this effect by trying multiple random restarts, for an n-start-LM algorithm. Localization performance of this levels out quickly, with the point of diminishing returns at about $n = 4$.

These results motivated us to construct a hybrid system, in which the MLP's output is used as the initial guess for LM. As we shall see in the next section, this MLP-start-LM performs very well indeed.

5 Comparative Performance: MLP, LM, MLP-start-LM

We tuned LM for good performance. We settled on LM with four re-starts at the fixed initial points $(0., -6.9282, 1.)$, $(-6., 3.4641, 1.)$, $(6., 3.4641, 1.)$, and $(0.01, 0.01, 6.1962)$, in units of cm. The covariance matrix was calculated from actual noise. We call this tuned system "4-start-LM." Modestly increasing the number of restarts increases the computational burden without much decreasing localization error.

Each of the MLP localizers from Section 3 was used as an LM initializer, for four variant MLP-start-LM localizers. The performance of all three localization systems, trained with various sorts of noise, is shown as a function of S/N in Figure 4. MLP-start-LM shows

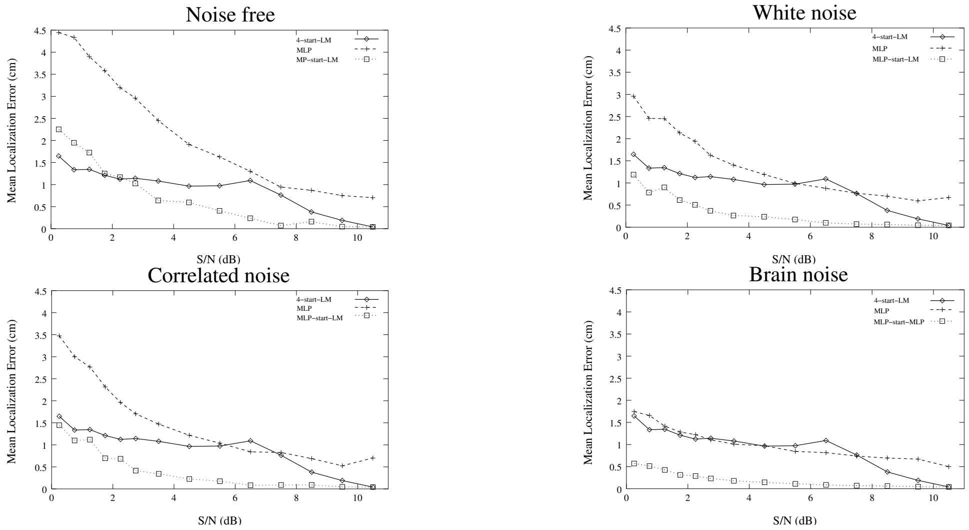


Figure 4: Mean localization error versus S/N for 4-start-LM, MLP, and MLP-start-LM. Four different sorts of training noise are shown. In all cases, testing used actual noise.

algorithm	4-start-LM	MLP				MLP-start-LM			
trained noise	—	N	W	C	B	N	W	C	B
time (ms)	**448.6**	0.5	0.5	0.5	**0.5**	57.4	42.1	45.6	**34.6**
error (mm)	**11.6**	29.2	17.7	19.5	**11.9**	10.9	5.0	6.1	**2.8**

Table 1: Comparison of performance on actual-noise test set of conventional Levenberg-Marquardt source localizer; trained MLP; and hybrid system. Each number is an average over 4,500 localizations, so the error bars are negligible. The training used various sorts of noise (N=None, W=White, C=Correlated, B=Actual). Naturally performance is best when the training noise is drawn from the same distribution as the testing noise.

the best localization performance across a broad range of training noise and S/N. A grand summary, averaged across various S/N conditions, is shown in Table 1.

6 Conclusion

We showed that initial guess is very important for the Levenberg-Marquardt localization method, and that LM performs much better with a good initial guess. The multilayer perceptron was shown to give good performance with reasonable accuracy across a range of mismatches between training and testing noise.

The MLP's localization accuracy was comparable to 4-start-LM's, at one five hundredth the computational burden. This motivated us to construct a hybrid system, MLP-start-LM, which improves the localization accuracy beyond any other practical techniques available to us (by a factor of about four) while reducing the computational burden to less than a tenth that of 4-start-LM.[1]

A number of extensions are planned in the immediate future: we will integrate a more sophisticated forward model already developed in our laboratory, we will experiment with

[1]The reason the computational burden is reduced by more than a factor of four, even though there is one LM instead of four, is that the initial guess is closer so the LM optimization is much faster.

secondary dipoles as the noise, and we will do a post-LM cleanup pass with an MLP trained to remove the bias that LM can introduce.

Acknowledgements

Supported in part by NSF CAREER award 97-02-311, the National Foundation for Functional Brain Imaging, an equipment grant from Intel corporation, and a gift from the NEC Research Institute.

References

Abeyratne, U. R., Kinouchi, Y., Oki, H., Okada, J., Shichijo, F., and Matsumoto, K. (1991). Artificial neural networks for source localization in the human brain. *Brain Topography*, 4:3–21.

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

Hoey, G. V., Clercq, J. D., Vanrumste, B., de Walle, R. V., Lemahieu, I., D'Havé, M., and Boon, P. (2000). Eeg dipole source localization using artificial neural networks. *Phys. Med. Biol.*, 45:997–1011.

Kinouchi, Y., Ohara, G., Nagashino, H., Soga, T., Shichijo, F., and Matsumoto, K. (1996). Dipole source localization of meg by bp neural networks. *Brain Topography*, 8:317–321.

Kwon, H. C., Lee, Y. H., Jun, S. C., Kim, J. M., Park, J. C., and Kuriki, S. (2000). Localization errors with 40-channel tangential fields. In Nenonen, J., Ilmoniemi, R., and Katila, T., editors, *Biomag2000 Proc. 12th Int. Conf. on Biomagnetism*, pages 943–946.

LeCun, Y., Kanter, I., and Solla, S. A. (1991). Second order properties of error surfaces: Learning time and generalization. In *Advances in Neural Information Processing Systems 3*, pages 918–924. Morgan Kaufmann.

Levenberg, K. (1944). A method for the solution of certain problems in least squares. *Quart. Appl. Math*, 2:164–168.

Lütkenhöner, B. (1994). Magnetic field arising from current dipoles randomly distributed in a homogeneous spherical volume conductor. *J. Appl. Phys.*, 75:7204–7210.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math*, 11:431–441.

Mosher, J. C., Leahy, R. M., and Lewis, P. S. (1999). Eeg and meg: Forward solutions for inverse methods. *IEEE Trans. Biomed. Engin.*, 46:245–259.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E., McClelland, J. L., and the PDP research group., editors, *Parallel distributed processing: Explorations in the microstructure of cognition, Volume 1: Foundations*. MIT Press.

Sarvas, J. (1987). Basic mathematical and electromaagnetic concepts of the biomagnetic inverse problem. *Phys. Med. Biol.*, 32:11–22.

Sun, M. and Sclabassi, R. J. (2000). The forward eeg solutions can be computed using artificial neural networks. *IEEE Trans. Biomed. Engin.*, 47:1044–1050.

Tang, A. C., Pearlmutter, B. A., and Zibulevsky, M. (1999). Blind separation of multichannel neuromagnetic responses. In *Computational Neuroscience*, pages 1115–1120. Published in a special issue of *Neurocomputing* volume 32–33 (2000).

Tang, A. C., Phung, D., Pearlmutter, B. A., and Christner, R. (2000). Localization of independent components from magnetoencephalography. In *International Workshop on Independent Component Analysis and Blind Signal Separation*, Helsinki, Finland.

Vigário, R., Jousmäki, V., Hämäläinen, M., Hari, R., and Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in Neural Information Processing Systems 10*. MIT Press.

Yuasa, M., Zhang, Q., Nagashino, H., and Kinouchi, Y. (1998). Eeg source localization for two dipoles by neural networks. In *Proc. 20th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 2190–2192.

Zhang, Q., Yuasa, M., Nagashino, H., and Kinouchi, Y. (1998). Single dipole source localization from conventional eeg using bp neural networks. In *Proc. 20th Ann. Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, volume 4, pages 2163–2166.