GROUPED L0 LEAST SQUARES PENALISED MAGNETOENCEPHALOGRAPHY

Ben Cassidy†, Victor Solo*†, Akila J. Seneviratne**

*School of Electrical Engineering and Telecommunications
University of New South Wales, Sydney, Australia
†Neuroscience Research Australia, Sydney, Australia
email: b.cassidy@unsw.edu.au

ABSTRACT

The distributed source inverse problem for Magnetoencephalography (MEG) requires regularisation in order to calculate a stable and accurate solution. Recently there has been interest in source reconstruction via the l_1 penalised least squares problem. Although this gives some sparseness the ideal measure for sparsity is the l_0 norm.

Here we develop a cyclic descent algorithm to solve the grouped-variable l_0 penalised least squares problem for an underdetermined linear system.

We demonstrate the utility of this method on real MEG data and show an increase in sparsity achieved using the l_0 instead of the l_1 penalty.

1. INTRODUCTION

The Magnetoencephalography (MEG) source localisation inverse problem involves estimating source positions of neural activation via external sampling of biomagnetic fields. Here we study the distributed source model which estimates amplitudes for a dense grid of current dipoles arranged on the cortical surface or throughout the brain volume at points determined from a structural MRI scan.

The model $y_t = Gq_t + \epsilon_t$ gives the observed field vector y_t as a linear sum of the source dipoles q_t representing focal neural activation; $\epsilon_t \sim N(0, \Sigma)$. $G_{n\times p}$ is the lead field matrix. The inverse problem is severely ill-conditioned for two reasons: Obtaining sufficient spatial resolution necessitates a highly sampled source space within cortex, so $p \gg n$. Thus the model is underdetermined and many possible source dipole arrangements will produce the same noise free magnetic field measurements. Secondly, the physics coupled with the relative distance of source and sensor locations causes the singular value profile of G to have many fewer than n significant singular values.

Hence, getting a unique and stable solution requires constraints to be placed upon the source distribution.

This work was partly supported by NIH P41 grant NCRR 2P41RR014075-11 and an ARC grant.

A standard approach to regularisation is to consider a penalised least squares problem

$$\min_q J(q) : \; J(q) = \frac{1}{2}\|y - Gq\|_2^2 + h\phi(q)$$

which compromises between minimising the residual energy of the linear model fit and penalising some charcteristic of the underlying source distribution $\phi(q)$. h is a penalty parameter. A common approach is to regularise via a weighted Tikhonov method, $\phi(q) = \|Wq\|_2^2$ where W is a diagonal weight matrix, which has the advantages of a robust solution and closed-form calculation. The noise-free version of this solves $\min \|Wq\|_2^2$ such that $y = Gq$ and is known as the Minimum Norm Estimate [1]. However this method promotes a smeared, non-zero estimate of cortical activation across the entire source space; generating a spatial map of estimated neural activation then requires applying a threshold not defined within the model.

Due to functional specialisation in brain regions we expect a spatially sparse solution for q_t. Applying a subquadratic penalty can give sparse solutions consistent with physiological understanding that macroscopic neural activation occurs in spatially clustered but focal regions.

An ideal sparse model would constrain the l_0 'norm' of the current distribution, i.e. the number of non-zero source dipoles. However l_0 regularisation is nonpolynomial (NP) hard: $J(q)$ becomes a non-convex function and a global minimum cannot be guaranteed when converging except by exhaustive search. So instead existing methods often approximate the sparse variable-choosing behaviour of l_0 by using an l_1 penalty to the least squares regression problem, as studied by [2] and popularly known as the LASSO [3]. In MEG the noise-free version of this is known as the Minimum Current Estimate (MCE) method [4]. MCE constrains the l_1 norm as $\phi(q) = \sum_i w_i |q_i|$, $\phi(q)$ being the total magnitude of the current distribution with predetermined source current directions. This problem is convex and can be solved using Linear Programming [5] or other methods such as cone programming [6].

Other related l_1 methods for MEG allow free-orientation of the estimated current dipoles by grouping the dipole com-

ponents $q_j = \sqrt{\left(q_{j,x}^2 + q_{j,y}^2 + q_{j,z}^2\right)}$ (independently developed for EEG by [7]) then constraining the sum of l_2 norms of each current dipole q_j as the previous l_1 problem. This takes the form of the group LASSO model [8] first proposed by [9]. Some models [6, 10, 11] extend to account for temporal smoothness, e.g. using an l_2 or other composite norm, which we do not consider here.

It is well understood that the l_0 penalty generates greater levels of sparsity than l_1. It can be shown that under certain restrictive conditions, such as a noise-free environment, that l_1 regularisation can give equivalent results as the l_0 [12]. This issue of equivalence has not been properly explored in the case of noisy data. While the spatial l_1 constraint on MEG dipole amplitude does promote some sparsity it does not provide equivalent solutions to an l_0 penalty as shown in this paper.

Practical differences between l_0 and l_1 norms for this problem are observed as scaling effects on coefficients. An l_0 norm causes immediate rejection of small coefficients; this differs from l_1 which shrinks all coefficients in the model by varying amounts and only discards coefficients as a result of this decrease.

In this paper we develop and demonstrate a grouped-variable l_0 regularised least squares cyclic descent algorithm (which we denote gL0LS-CD) suitable for solving the MEG inverse problem. This method has the advantage of the l_1-equivalent group-LASSO in that it penalises groups of variables to be identically zero, which corresponds to grouping dipole direction components in the MEG case. It is an extension of the scalar l_0 penalised multivariate regression cyclic descent based method presented in [13].

In Section 2 we derive the gL0LS-CD algorithm and provide a simple implementation. In Section 3 we compare the results of the l_0 penalised gL0LS-CD and the l_1 penalised gLASSO algorithms on real MEG data. Conclusions are offered in Section 4.

2. METHOD

Here we develop a grouped-l_0 penalised least squares cyclic descent (gL0LS-CD) algorithm for calculating a regularised inverse of an underdetermined linear system.

For a linear model $y = X\beta + \epsilon$, arrange the columns of X into m groups, $[X_u]_{n\times p_u}$ of size p_u, $u = 1, \ldots, m$ so $\sum_1^m p_u = p$. β_u is the corresponding p_u-vector of coefficients.

Assume group-wise column scaling so that $X_u^T X_u = I_{p_u}$.

Consider cyclic descent minimization of

$$J(\beta) = \|y - X\beta\|^2 + h\sum_{v=1}^{m} I(\|\beta_v\| \neq 0)$$

where $I(\cdot)$ is the indicator and so $\sum_{v=1}^m I(\|\beta_v\| \neq 0)$ is the number of non-zero coefficient groups. Weighting parameter h determines the emphasis between minimising either the error variance or the sparsity of the solution.

Theorem 1 To optimise at the $(k+1)$th iteration we update for m steps, through each group of parameters β_v, $v = 1, \ldots, m$. At each step to update β^k,

1. Compute $e^k = y - X\beta^k$
2. Compute $\gamma_v^k = X_v^T e^k$
3. $\beta_v^{k+1} = \left(\gamma_v^k + \beta_v^k\right) I\left(\left\|\gamma_v^k + \beta_v^k\right\| \geq \sqrt{h}\right)$

Then after iterating this for $v = 1, \ldots, m$ we have updated β^k to β^{k+1}.

Derivation Fix all coefficient groups β_u at their values at the kth iteration except for β_v. Then

$$J = \left\|e_{-v}^k - X_v\beta_v\right\|^2 + h\sum_{u\neq v} I_u^k + hI_v$$

where $I_u^k = I\left(\left\|\beta_u^k\right\| \neq 0\right)$; $e_{-v}^k = y - \sum_{u\neq v} X_u\beta_u^k$.

Expanding gives

$$J = \left\|e_{-v}^k\right\|^2 + h\sum_{u\neq v} I_u^k - 2e_{-v}^{k,T} X_v\beta_v + \|\beta_v\|^2 + hI_v$$

now set

$$z_v^{k,T} = e_{-v}^{k,T} X_v$$

and add and subtract

$$z_v^{k,T} z_v^k$$

to get

$$\begin{aligned} J &= J_{-v} + R(\beta_v) \\ J_{-v} &= \left\|e_{-v}^k\right\|^2 - \left\|z_v^k\right\|^2 + h\sum_{u\neq v} I_u^k \\ R(\beta_v) &= \left\|z_v^k - \beta_v\right\|^2 + hI_v \end{aligned}$$

Now J_{-v} does not depend on β_v and so can be dropped.

Lemma The minimizer of $R(\beta_v)$ is

$$\beta_v^{k+1} = z_v^k I\left(\left\|z_v^k\right\| \geq \sqrt{h}\right)$$

Proof. $R(0) = \left\|z_v^k\right\|^2$, otherwise $R(\beta_v)$ is minimised at $\beta_v = z_v^k$ with corresponding value h. Thus the minimum is at 0 if $\left\|z_v^k\right\|^2 \leq h \equiv \left\|z_v^k\right\| \leq \sqrt{h}$ and the result follows.

We can rewrite

$$\begin{aligned} z_v^k &= X_v^T e_{-v}^k \\ &= X_v^T\left(e^k + X_v\beta_v^k\right) \\ &= \gamma_v^k + \beta_v^k \\ e^k &= y - X\beta^k \\ \gamma_v^k &= X_v^T e^k \end{aligned}$$

then the gL0LS-CD solution can be written

$$\beta_v^{k+1} = \left(\gamma_v^k + \beta_v^k\right) I\left(\left\|\gamma_v^k + \beta_v^k\right\| \geq \sqrt{h}\right) \quad (1)$$

3. RESULTS

For MEG analysis this algorithm is applied separately at each time point as a 'snapshot' analysis. For the model

$$y = Gq + \epsilon$$

we operate on mean-corrected data and lead-field matrix. The model is pre-whitened as $G = \Sigma^{-\frac{1}{2}} G_{\text{raw}}$, $y = \Sigma^{-\frac{1}{2}} y_{\text{raw}}$. The columns of the lead-field matrix G are scaled via a depth-weighting procedure [14] before orthonormalising the column groups using a Cholesky decomposition as follows,

$$\begin{aligned} Gq &= \sum_{u=1}^{m} g_u q_u \quad [q_u \text{ are 3-vectors}] \\ &= g_u \left(g_u^T g_u\right)^{-\frac{1}{2}} \left(g_u^T g_u\right)^{\frac{1}{2}} q_u \\ &= \tilde{g}_u \tilde{q}_u \end{aligned}$$

where $\tilde{g}_u = g_u \left(g_u^T g_u\right)^{-\frac{1}{2}}$.

The inverse problem is therefore solved for parameters $\tilde{q}_u = \left(g_u^T g_u\right)^{\frac{1}{2}} q_u$ and then unscaled to calculate q. The Minimum Norm Estimate is used as a seed solution for the subsequent optimisation to increase the likelihood of the l_0-descent algorithm converging to a global minimum.

We compare results from real data using the grouped-l_1 (gLASSO) and grouped-l_0 (gL0LS-CD) penalty algorithms. The data is from a basic auditory stimulus experiment and is taken as a time snapshot from averaged stimulus measurements. MEG data is sampled at 1000Hz by 157 coaxial first-order gradiometers (KIT, Kanazawa, Japan); $\dim(G) = 157 \times 24576$.

The calculated activation pattern from a single time snapshot using the gL0LS-CD algorithm gives a focal estimate of source dipole activity in the secondary auditory cortex (Fig. 1). This activation is detected on the upper temporal gyrus, immediately adjacent to the primary auditory cortex on the supratemporal gyrus. Although this estimate does not correspond to the expected region of the primary auditory cortex, the small spatial error could be attributed to lack of orientation constraint on the source estimates.

Choosing the regularisation weighting factor h is vital to achieve plausible results. The l_0 regularised solution shown in Fig. 1 is stable within a reasonable range; small h gives a distributed dipole estimate whereas larger h gives a null result.

Results from the l_1 penalised gLASSO algorithm in Fig. 2 also show focal dipole estimates. The gLASSO results are more distributed than from the gL0LS-CD approach and almost certainly include false positive dipole activation estimates, since the detected regions extend into parietal lobe and middle temporal gyrus which are not expected to be involved in the simple auditory experiment task.

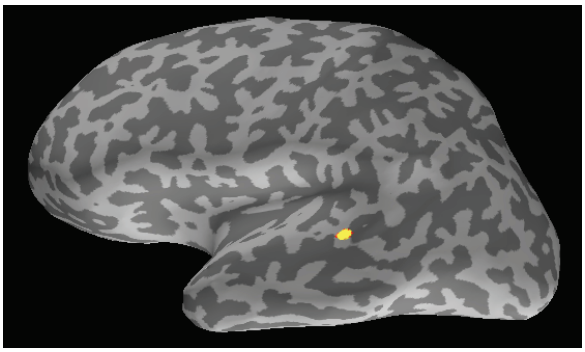


Fig. 1. MEG activation map from auditory stimulus calculated using a grouped-l_0 regularised least squares algorithm.

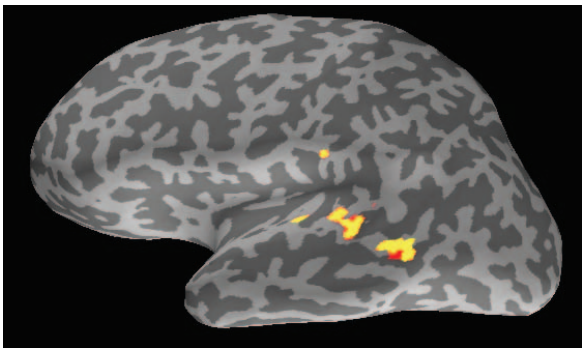


Fig. 2. Map of activation from auditory stimulus calculated using a grouped-l_1 regularised least squares (gLASSO) algorithm.

Increasing the regularisation parameter h causes the l_1 solution to converge towards the posterior lower temporal gyrus leading to an obviously false-positive result.

These are preliminary but encouraging results for this new approach. Including other model assumptions incorporating temporal and further spatial distributions is well known to give improved estimates of dipole amplitude and we will pursue these issues in further work. In this work we are primarily concerned with the improved level of sparsity gained by using an l_0 penalty instead of the l_1. The results shown here demonstrate that increased sparsity.

4. CONCLUSION

In this work we have developed an algorithm to solve the grouped-variable l_0-regularised least squares regression problem using cyclic descent. Applied to the Magnetoencephalography inverse problem, the algorithm constrains the three components of the source vector at each voxel to be jointly zero or jointly non-zero.

We have demonstrated the utility of this algorithm compared to an l_1 approach in a comparative study on real MEG data to show spatially focal source dipole estimates.

Future work on this topic will extend the model to account for piece-wise temporal smoothness within the data. There is also scope to develop more rigorous selection criteria for the l_0 regularisation parameter equivalent to the grouped-l_1 problem discussed in [15].

5. REFERENCES

[1] M. Hämäläinen and R. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Medical and Biological Engineering and Computing*, vol. 32, no. 1, pp. 35–42, 1994.

[2] S. Alliney and S. Ruzinsky, "An algorithm for the minimization of mixed l (1) and l (2) norms with application to bayesian estimation," *IEEE Trans. Sig. Proc.*, vol. 42, no. 3, pp. 618–627, 1994.

[3] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Jl. Roy. Stat. Soc. B*, pp. 267–288, 1996.

[4] K. Uutela, M. Hämäläinen, and E. Somersalo, "Visualization of magnetoencephalographic data using minimum current estimates," *Neuroimage*, vol. 10, no. 2, pp. 173–180, 1999.

[5] K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," *IEEE Trans. Biomed. E.*, vol. 42, no. 6, pp. 608–615, 1995.

[6] W. Ou, M. Hämäläinen, and P. Golland, "A distributed spatio-temporal EEG/MEG inverse solver," *NeuroImage*, vol. 44, no. 3, pp. 932–946, 2009.

[7] L. Ding and B. He, "Sparse source imaging in electroencephalography with accurate field modeling," *Human brain mapping*, vol. 29, no. 9, pp. 1053–1067, 2008.

[8] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Jl. Roy. Stat. Soc. B*, vol. 68, no. 1, pp. 49–67, 2006.

[9] S. Bakin, *Adaptive regression and model selection in data mining problems*. PhD thesis, School of Mathematical Sciences, The Australian National University, Canberra, 1999.

[10] K. Friston, L. Harrison, J. Daunizeau, S. Kiebel, C. Phillips, N. Trujillo-Barreto, R. Henson, G. Flandin, and J. Mattout, "Multiple sparse priors for the M/EEG inverse problem," *NeuroImage*, vol. 39, no. 3, pp. 1104–1120, 2008.

[11] A. Gramfort, D. Strohmeier, J. Haueisen, M. Hamalainen, and M. Kowalski, "Functional brain imaging with m/eeg using structured sparsity in time-frequency dictionaries," in *Information Processing in Medical Imaging*, pp. 600–611, Springer, 2011.

[12] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, p. 2197, 2003.

[13] A. Seneviratne and V. Solo, "On vector l0 penalized multivariate regression," in *Proc IEEE ICASSP, to appear*, 2012.

[14] F. Lin, T. Witzel, S. Ahlfors, S. Stufflebeam, J. Belliveau, and M. Hamalainen, "Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates," *Neuroimage*, vol. 31, no. 1, pp. 160–171, 2006.

[15] V. Solo and M. Ulfarsson, "Threshold selection for group sparsity," in *Proc. IEEE ICASSP 2010*, pp. 3754–3757, IEEE, 2010.