

New Solver to the Ill-posed Problem in Magnetoencephalography

Norikazu SAITO (齊藤宣一)

Department of Mathematics, Faculty of Education, Toyama University

Takashi SUZUKI (鈴木貴)

Department of Mathematics, Graduate School of Science, Osaka University

Yoshiaki ADACHI (足立善昭)

Applied Electronics Laboratory, Kanazawa Institute of Technology

Hisashi KADO (賀戸久)

Applied Electronics Laboratory, Kanazawa Institute of Technology

Masahiro SIMOGAWARA (下川原正博)

Applied Electronics Laboratory, Kanazawa Institute of Technology

This article proposes a new scheme for MEG data analysis. For readers who are not familiar with MEG, we have provided preliminary descriptions on MEG analysis and discrete inverse problems. If MEG is familiar, they can skip over to section 1 directly.

MEG Analysis

Magnetic fields are measured from several parts of the human body; heart, liver, and lungs. But the brain also radiates them because of the neuron currents inside as a result of its activities. They are rather weak, but the recent progress of device, called SQUID (superconducting quantum interference device) following its mechanism, makes it possible to measure them. It is based on the quantization of magnetic flux densities and Josephson tunneling, both of which occur in the range of low temperatures. Now they are being applied to cerebral physiology, psychology, linguistics, clinical medicine, and so forth. (See Ilmoniemi [11].)

More than 100 SQUID sensors, called *channels*, are arranged in a helmet-shape array to cover the human head. At each channel, one component of the

magnetic field is measured. As a result of the activities of neurons, magnetic fields arise outside the brain and are taken impulsively as data. On the other hand, detailed shape of the brain of the testee is investigated in advance in use of MRI (magnetic resonance image). Then, the data analysis is made to decide locations and moments of neuron currents inside. Such a research is called the *magnetoencephalography*, or *MEG* in short. Theoretical foundation was given by Sarvas [13] based on the integral equation due to Geselowitz [5] and Geselowitz and Grynszpan [6].

Direct problem is formulated as follows. First, Maxwell's equation controls the total current density J and the magnetic field B in such a way as

$$\nabla \cdot B = 0 \quad \text{and} \quad \nabla \times B = \mu_0 J. \quad (1)$$

Here, the state is supposed to be quasi-statistic and permeability inside the brain is put to be equal to that in the vacuum. We have $\mu_0 = 4\pi/c$ with the velocity c of light. The operation ∇ denotes the gradient and the inner and the outer products in $\mathbf{R}^3$ are written as $\cdot$ and $\times$, respectively.

Total current density has the form

$$J = J^p - \sigma(x)\nabla V$$

with $J^p(x)$ and $E = -\nabla V(x)$ standing for the neuron current mentioned above and the electric field caused by it, respectively. Sometimes J^p is called the *primary* current, and $-\sigma(x)\nabla V$ the *secondary* one. The brain in consideration is denoted by a bounded domain $\Omega \subset \mathbf{R}^3$ with smooth boundary $\partial\Omega$. We suppose that $J^p(x)$ is a smooth vector field on $\overline{\Omega}$, with null normal component on $\partial\Omega$ and put $J^p(x) = 0$ outside $\overline{\Omega}$. The conductivity is supposed to have the form

$$\sigma(x) = \begin{cases} \sigma_I & (x \in \Omega) \\ \sigma_O & (x \in \overline{\Omega}^c) \end{cases},$$

where σ_{IO} are non-negative constants, and $V(x)$ is a piecewise smooth continuous function in $\mathbf{R}^3$.

In 1970, Geselowitz [5] asserted that then

$$B(x) = \frac{\mu_0}{4\pi} \int_{\Omega} J^p(y) \times \frac{x-y}{|x-y|^3} dy - \frac{\mu_0}{4\pi} (\sigma_I - \sigma_O) \int_{\partial\Omega} V(y) n_y \times \frac{x-y}{|x-y|^3} dS_y \quad (2)$$

follows for $x \in \overline{\Omega}^c$, where the unit normal vector $n = n_y$ is taken to be outer from Ω . Also Geselowitz and Grynszpan [6] said that if Ω is a ball and

$\sigma_O = 0$, then it holds that

$$\begin{aligned} B(x) &= \mu_0 \nabla U(x) \\ U(x) &= \frac{1}{4\pi} \int_{\Omega} \frac{J^P(y) \times y}{|x - y| (|x| |x - y| + x \cdot (x - y))} dy \cdot x \end{aligned} \quad (3)$$

for $x \in \overline{\Omega}^c$. Equality (2) is called the *Geselowitz equation*, and system (3) the *spherical model*.

The MEG analysis of [13] assumes the spherical model and also the primary current $J^P(x)$ to be a combination of dipoles such as

$$J^P(x) = \sum_{i=1}^m Q_i \delta(x - a_i), \quad (4)$$

where $a_i \in \Omega$ and $Q_i \in \mathbf{R}^3$ denote the position and the moment of a dipole, respectively. Then (3) is written as

$$U(x) = \frac{1}{4\pi} \sum_{i=1}^m \frac{Q_i \times a_i \cdot x}{|x - a_i| (|x| |x - a_i| + x \cdot (x - a_i))}. \quad (5)$$

Unknown parameters $\{a_i, Q_i\}$ are so determined as to adjust with discretely measured values of $\{n \cdot B\}$ on $\partial\Omega$ in way of the method of least square approximation denoted by *LSA*.

Relation (4) is called the *dipole hypothesis*. Examining its validity is beyond mathematics, but is widely accepted as a reaction of the brain to a limited number of stimulations. Actually, a single dipole fits the real data of magnetic fields taken under one stimulation quite well. More precisely, if one takes a sphere close to the brain based on MRI analysis and gives periodic stimulations to one point of the testee's body, then the observed data are traced very precisely by the assumed magnetic field $B(x) = \mu_0 \nabla U(x)$, where $U(x)$ has the form (5) with $N = 1$. (See [11].) It fits also the functional brain mapping by Penfield and others. Since the system is supposed to be linear, super-position of dipoles as in (4) will be realistic for more general situations. It is of course a rough model of the functional brain. For example, with latencies of roughly 100ms after median nerve stimulation, several distinct brain areas can be simultaneously active.

However, relation between the numbers of assumed dipoles and that of channels leads LSA to be under- or over- determined exclusively, of which difficulties are essentially different. Furthermore, the sphere may not be a

good approximation of the shape of head. In this connection, Fokas and Kurylev [4] and Fokas, Gel'fand, and Kurylev [3] made use of the method of regularization instead of assuming dipole hypothesis. See also Hämäinen and Ilmoniemi [7] and Dale and Sereno [2]. On the other hand, it has been reported that the global least-square solutions of multi-dipole models are obtained with suitable methods. See, for example, [17] and [10].

Discrete Inverse Problems

One can summarize that the *discrete inverse problem* takes regards to the finiteness and the error in the observing process, while the unknown are restricted to a finite number of parameters. In the context of MEG, if the number of prescribed dipoles and that of observed data are balanced, then it becomes generically well-posed. However, even under the dipole hypothesis, the number of dipoles is preferably to be left unknown. The data observed at many channels are also to be used efficiently. Here, we develop a general theory of the discrete inverse problem from those points of view.

Let $\varphi : \mathbf{R}^n \rightarrow \mathbf{R}^m$ be a nonlinear $C^{2,1}$ mapping, representing

$$\text{physical law} \times \text{observation}$$

and $z \in \mathbf{R}^m$ the measured value with errors permitted. Then the discrete inverse problem is so formulated as to find $x \in \mathbf{R}^n$ satisfying $\varphi(x) = z$, of which solution is called the *strict solution*. Note that it is over-determined and under-determined according to $n < m$ and $n > m$, respectively.

In the over-determined case, the strict solution is hard to obtain; the discrete inverse problem is reformulated as the least square problem to find $x \in \mathbf{R}^n$ satisfying $\mathcal{J}(x) = j$, where $\mathcal{J}(x) = \frac{1}{2} \|\varphi(x) - z\|^2$ and $j = \inf_{\mathbf{R}^n} \mathcal{J}$. Its solution is called the *least square solution*, but practically one can examine its local minimality only. Any local minimum is regarded as a least square solution and sometimes is not unique. If $\mathcal{J}(x)$ is satisfactorily small, one says that x has a *high accuracy*. A particular least square solution has to be selected, regarded with a priori informations and accuracies. We say that the least square problem is *quasi-identifiable* if any local minimum of $\mathcal{J}$ is isolated. This allows one to compute and select one of them as a desirable solution.

A criterion for quasi-identifiability is high-accuracy and rank condition. The latter means that $\varphi'(x_0)^* \varphi'(x_0) : \mathbf{R}^n \rightarrow \mathbf{R}^n$ is non-singular, where $*$ denotes the transpose of the matrix. More precisely, the existence of $x_0 \in \mathbf{R}^n$

provided with the properties of high-accuracy and rank condition assures the quasi-identifiability in its neighbourhood. Because the rank condition is generically holds by $n < m$, only high-accuracy is to be examined in the over-determined case. If high-accuracy is achieved in MEG data analysis for many channels and very few prescribed dipoles, then it assures the validity of dipole hypothesis mathematically.

We have the following theorem, where $B_r(x_0) = \{x \in \mathbf{R}^n \mid |x - x_0| < r\}$, $\|G\|_{L^\infty(B_r(x_0))} = \sup_{x \in B_r(x_0)} |G(x)|$, and

$$[G]_{Lip(B_r(x_0))} = \sup_{x \neq y, x, y \in B_r(x_0)} \frac{|G(x) - G(y)|}{|x - y|}.$$

Theorem 1 *Given $C_0, C_1 > 0$, we have $r > 0$ and $\delta > 0$ such that if there is $x_0 \in \mathbf{R}^n$ satisfying $\sup_{x \in B_{3r}(x_0)} \|[\varphi'(x)^* \varphi'(x)]^{-1}\| \leq C_0$,*

$$\|\varphi'\|_{L^\infty(B_{2r}(x_0))} + \|\varphi''\|_{L^\infty(B_{3r}(x_0))} + [\varphi'']_{Lip(B_{2r}(x_0))} \leq C_1,$$

and $\sup_{x \in B_{3r}(x_0)} \|\varphi(x) - z\| < \delta$, then $\inf_{B_r(x_0)} \mathcal{J}$ is attained by a unique element of $B_r(x_0)$.

Proof: We have

$$\left. \frac{d}{dt} \mathcal{J}(x + ty) \right|_{t=0} = \langle \varphi'(x)y, \varphi(x) - z \rangle$$

and hence $x \in \mathbf{R}^n$ is a critical point of $\mathcal{J}$ if and only if

$$\varphi'(x)^* \varphi(x) = \varphi'(x)^* z \quad (6)$$

holds. We also have

$$\begin{aligned} \left. \frac{d^2}{dt^2} \mathcal{J}(x + ty) \right|_{t=0} &= \langle [\varphi''(x)y]y, \varphi(x) - z \rangle + \langle \varphi'(x)y, \varphi'(x)y \rangle \\ &\geq (C_0^{-1} - \delta C_1) |y|^2 \end{aligned} \quad (7)$$

for $x \in B_{3r}(x_0)$ and hence any critical point of $\mathcal{J}$ in $B_{3r}(x_0)$ is a non-degenerate local minimum in the case of $\delta \in (0, C_0^{-1}C_1^{-1})$. We show that $x \in \overline{B_{2r}(x_0)}$ satisfying (6) exists uniquely and is contained in $B_r(x_0)$ if $\delta, r > 0$ are sufficiently small. Then the conclusion follows.

For this purpose, we note that the mean value theorem implies

$$\varphi(x) = \varphi(x_0) + \varphi'(x_0)(x - x_0) + R(x)$$

with

$$R(x) = \frac{1}{2} [\varphi''(x_1)(x - x_0)] (x - x_0),$$

where $x_1 \in [x, x_0]$. This gives

$$\|R\|_{L^\infty(B_{2r}(x_0))} \leq 2 \|\varphi''\|_{L^\infty(B_{2r}(x_0))} r^2$$

and

$$[R]_{Lip(B_{2r}(x_0))} \leq 2 [\varphi'']_{Lip(B_{2r}(x_0))} r^2 + 2 \|\varphi''\|_{L^\infty(B_{2r}(x_0))} r$$

by a simple calculation. Similarly, we have

$$\varphi'(x)^* = \varphi'(x_0)^* + E$$

with

$$\|E\|_{L^\infty(B_{2r}(x_0))} \leq 2 \|\varphi''\|_{L^\infty(B_{2r}(x_0))} r$$

and

$$[E]_{Lip(B_{2r}(x_0))} \leq 2 [\varphi'']_{Lip(B_{2r}(x_0))} r + \|\varphi''\|_{L^\infty(B_{2r}(x_0))}.$$

We see that (6) is equivalent to $x = x_0 + \Phi(x)$, where

$$\begin{aligned} \Phi(x) &= [\varphi'(x_0)^* \varphi'(x_0)]^{-1} \\ &\quad \cdot (\varphi'(x)^* (x - \varphi(x_0)) - \varphi'(x)^* R - E \varphi'(x_0)(x - x_0)). \end{aligned}$$

Here we have

$$\|\Phi\|_{L^\infty(B_{2r}(x_0))} \leq C_0 C_1 (\delta + \|R\|_{L^\infty(B_{2r}(x_0))} + \|E\|_{L^\infty(B_{2r}(x_0))} 2r)$$

and

$$\begin{aligned} [\Phi]_{Lip(B_{2r}(x_0))} &\leq C_0 [\varphi']_{Lip(B_{2r}(x_0))} (\delta + \|R\|_{L^\infty(B_{2r}(x_0))}) \\ &\quad + C_0 \|\varphi'\|_{L^\infty(B_{2r}(x_0))} [R]_{Lip(B_{2r}(x_0))} \\ &\quad + C_0 C_1 ([E]_{Lip(B_{2r}(x_0))} 2r + \|E\|_{L^\infty(B_{2r}(x_0))}). \end{aligned}$$

Therefore, if $\delta, r > 0$ are small, Φ is a contraction mapping on $\overline{B_{2r}(x_0)}$. Furthermore, $\Phi(\overline{B_{2r}(x_0)}) \subset B_r(x_0)$ is achieved, and the proof is complete.

1 Introduction

Usual MEG data analysis in use of SQUID is based on Sarvas [13]. It adopts the spherical model of Geselowitz and Grynszpan [6]. If the state is quasi-static and the permeability inside the brain is equal to that in vacuum, then Maxwell's equation controls the total current density J and the magnetic field B in such a way as

$$\nabla \cdot B = 0 \quad \text{and} \quad \nabla \times B = \mu_0 J, \quad (8)$$

where $\mu_0 = 4\pi/c$ with the velocity c of light. Here, J has the form

$$J = J^p - \sigma(x)\nabla V$$

with $J^p(x)$ and $E = -\nabla V(x)$ standing for the neuron current and the electric field caused by it, respectively. Conductivity is supposed to be

$$\sigma(x) = \begin{cases} \sigma_I & (x \in \Omega) \\ \sigma_O & (x \in \bar{\Omega}^c) \end{cases},$$

where σ_{IO} are non-negative constants. Then, in way of the Geselowitz equation [5], it follows that

$$\begin{aligned} B(x) &= \mu_0 \nabla U(x) \\ U(x) &= \frac{1}{4\pi} \int_{\Omega} \frac{J^p(y) \times y}{|x-y| (|x| |x-y| + x \cdot (x-y))} dy \cdot x \end{aligned}$$

for $x \in \bar{\Omega}^c$, if Ω is a ball and $\sigma_O = 0$. Sarvas [13] assumes also the dipole hypothesis so that the primary current $J^p(x)$ is put to be

$$J^p(x) = \sum_{i=1}^m Q_i \delta(x - a_i), \quad (9)$$

where $a_i \in \Omega$ and $Q_i \in \mathbf{R}^3$ denote the position and the moment of a dipole, respectively. This implies

$$U(x) = \frac{1}{4\pi} \sum_{i=1}^m \frac{Q_i \times a_i \cdot x}{|x - a_i| (|x| |x - a_i| + x \cdot (x - a_i))}$$

and unknown parameters $\{a_i, Q_i\}$ are so determined as to adjust with discretely measured values of $\{n \cdot B\}$ on $\partial\Omega$ through the method of least square

approximation, where n denotes the outer unit normal vector. Usually, the number of observed data is much more than that of prescribed dipoles and the problem formulated in this way is over-determined. If high-accuracy is achieved by a set of dipoles, then generically there is a unique local minimum near by it.

Our motivation lies in the following questions. First, ball may not be a good approximation of the shape of brain. In that case, how is the secondary current $-\sigma(x)\nabla V$ to be picked up? Second, what is to be done if the primary current can not be suspected to be one or two dipoles? This case occurs when the primary current is distributed in a wide area, or when the number of dipoles is not prescribed although it is assumed to be finite.

2 Current Element Distribution

To treat those problems, we observe that if $J(x)$ is a single dipole $Q_k\delta(x-a_k)$, then it produces the magnetic field

$$\frac{\mu_0 Q_k \times (x - a_k)}{4\pi |x - a_k|^3}.$$

Actually, we have

$$\nabla \times \left(\frac{Q \times (x - a)}{4\pi |x - a|^3} \right) = Q\delta(x - a) \quad \text{and} \quad \nabla \cdot \left(\frac{Q \times (x - a)}{4\pi |x - a|^3} \right) = 0$$

and hence $\frac{Q \times (x - a)}{4\pi |x - a|^3}$ is a fundamental solution. We regard $Q_k\delta(x - a_k)$ as a *current element*. A lot of elements distributed in Ω may be able to trace the total current density $J(x)$. On the other hand, if one takes the spherical model and ν_j is normal to $\partial\Omega$, then the data $B(x_j) \cdot \nu_j$'s produced by $J^p(x)$ and $J(x)$ are the same, because

$$\left[\nabla \cdot \left(\frac{Q \times a \cdot x}{|x - a| (|x| |x - a| + x \cdot (x - a))} \right) \right] \cdot n = \frac{Q \times (x - a)}{|x - a|^3} \cdot n$$

holds if $\Omega = \{x \in \mathbf{R}^3 \mid |x| < 1\}$. Therefore, if Ω is close to a ball and almost normal components of the magnetic field are observed, then one can also require for those elements to recover the primary current $J^p(x)$ in the form

Based on those observations, we propose the method of current element distribution in the following way. Let the number of channels, their positions, and the directions be M (around 100), $w_j \in \partial\Omega$, and $\nu_j \in \mathbf{S}^2$, respectively, where $1 \leq j \leq M$. Taking N sufficiently large (say, 200), one introduces the mapping

$$\varphi : (Q, a) \in \mathbf{R}^{6N} \mapsto \left(\sum_{k=1}^N \frac{\mu_0 Q_k \times (w_j - a_k)}{4\pi |w_j - a_k|^3} \cdot \nu_j \right)_{1 \leq j \leq M} \in \mathbf{R}^M$$

and the functional

$$\mathcal{J}(x) = \frac{1}{2} \|\varphi(x) - z\|_{\mathbf{R}^M}^2,$$

where $x = (Q, a)$, $Q = (Q_1, Q_2, \dots, Q_N) \in \mathbf{R}^{3N}$, $a = (a_1, a_2, \dots, a_N) \in \mathbf{R}^{3N}$, and $z = (z_1, z_2, \dots, z_M) \in \mathbf{R}^M$. The value $z_j \in \mathbf{R}$ stands for the observed datum of the ν_j component of the magnetic field $B(x)$ measured at j -th channel. First, N random current elements are distributed in Ω . One element is selected randomly and perturbed so that $\mathcal{J}$ decreases. Then the next one is chosen also randomly to continue the process. If $\mathcal{J}$ does not decrease, that element is left unperturbed. If the directions of two adjacent elements happen to be opposite, then randomly selected one of them is moved away. This procedure is continued until a local minimum of $\mathcal{J}$ is achieved.

Contrarily to the classical one, method of current element distribution is set to be under-determined. High accuracy is easy to achieve, but the solution is not unique. In fact, generically the set

$$\mathcal{M} = \{x = (Q, a) \in \mathbf{R}^{6N} \mid \varphi(x) = z\}$$

forms a manifold of $6N - M$ dimension. Thus, we have to re-formulate the problem as to find $x \in \mathcal{M}$ provided with the most desirable properties. Usually, this is done by another variational structure. One introduces a value function d on $\mathcal{M}$ and tries to find an element in $\mathcal{M}$ which attains $\sup_{\mathcal{M}} d$. However, introducing such d is not trivial, and the standard formulation does not seem to be definite in the data analysis of MEG. We propose a different method in this paper. The procedure is different according to the cases where the current element distribution is regarded as the total current density and where it is regarded as the neuron current composed of finite number of dipoles, respectively.

3 Parallel Optimization

Our scheme to solve under-determined problems is divided into three steps, *approaching*, *freezing*, and *melting*.

In approaching, the iterative sequence $\{x_\ell\} \subset \mathbf{R}^{6N}$ is so taken as to improve accuracies. For this procedure, one can apply (quasi-) Newton method. In this case, it is worth noting that the linearized operator is degenerate and the regularization process is efficient. In fact, we have

$$\left. \frac{d}{dt} \mathcal{J}(x + ty) \right|_{t=0} = \langle \varphi'(x)y, \varphi(x) - z \rangle$$

and hence $x \in \mathbf{R}^{6N}$ is a critical point of $\mathcal{J}$ if and only if

$$\varphi'(x)^* \varphi(x) = \varphi'(x)^* z \quad (10)$$

holds. We also have

$$\left. \frac{d^2}{dt^2} \mathcal{J}(x + ty) \right|_{t=0} = \langle \varphi''(x)[y, y], \varphi(x) - z \rangle + \langle \varphi'(x)y, \varphi'(x)y \rangle$$

and hence the linearized operator is given by $(\varphi(x) - z) \cdot \varphi''(x) + \varphi'(x)^* \varphi'(x)$. Because $\varphi : \mathbf{R}^{6N} \rightarrow \mathbf{R}^M$, its rank is less than or equal to $M < 6N$.

On the other hand, we have

$$\begin{aligned} \mathcal{J}(x_{\ell+1}) &= \mathcal{J}(x_\ell) + \mathcal{J}'(x_\ell) [\Delta x_{\ell+1}] + O(\|\Delta x_{\ell+1}\|^2) \\ &= \mathcal{J}(x_\ell) + (\varphi'(x_\ell) \Delta x_{\ell+1}, \varphi(x_\ell) - z) + O(\|\Delta x_{\ell+1}\|^2) \end{aligned} \quad (11)$$

with $\dim \text{Ker } \varphi'(x_\ell) = 6N - M$ generically, where $\Delta x_{\ell+1} = x_{\ell+1} - x_\ell$. Therefore, if one makes use of the gradient method, the perturbation is so taken as $\Delta x_{\ell+1} = s \cdot \varphi'(x_\ell)^* (z - \varphi(x_\ell))$ with

$$0 < s = o(\|\varphi'(x_\ell)^* (\varphi(x_\ell) - z)\|) = o(\|\varphi'(x_\ell)\| \cdot \mathcal{J}(x_\ell)^{1/2}).$$

Finally, if one adopts the method of the random perturbation in this process, then it is efficient to take $\Delta x_{\ell+1}$ in $\text{Ker } \varphi'(x_\ell)^\perp = \text{Ran } \varphi'(x_\ell)^*$ in view of (11). Actually, if one makes use of the singular decomposition of the matrix $\varphi'(x)^*$, this selection $\Delta x_{\ell+1} \in \text{Ker } \varphi'(x_\ell)^\perp = \text{Ran } \varphi'(x_\ell)^*$ is easy to realize. See Stoer and Bulirsch [15].

The next process is freezing. From (8) again, it is seen that the iterative sequence is hard to improve the accuracy in the region where

$$\|\varphi'(x_\ell) [\varphi(x_\ell) - z]\| \sim \|\Delta x_{\ell+1}\|.$$

This condition may be replaced by $\|\Delta x_{\ell+1}\| \sim \|\varphi'(x_\ell)\| \cdot \mathcal{J}(x_\ell)^{1/2}$, and we call this area the *freezing zone*. There, the iterative sequence looks like to have converged. As we shall describe later, such a phenomenon can happen in the area even far from $\mathcal{M}$, but this criterion can be used to examine whether the sequence is actually freezing or not. Another application of the freezing zone is to get higher accuracies. That is, even when x_ℓ comes into the freezing zone, one can still improve the accuracy by reducing $\|\Delta x_{\ell+1}\|$. However, this process cannot come into the region where SQUID recognizes the object, that is, over the spatial resolution.

That situation of freezing can be broken more efficiently by making use of the tangent space $\mathcal{T}_{x_\ell}\mathcal{M}$. This process is indicated as melting. In fact, one can remove the freezed sequence without losing accuracy so much by taking $\Delta x_{\ell+1} \in \mathcal{T}_{x_\ell}\mathcal{M}$. Here, we have $\mathcal{J}(x_\ell) \ll 1$ and hence $\mathcal{T}_{x_\ell}\mathcal{M} \approx \text{Ker } \varphi'(x_\ell)$. Therefore, we are able to replace the above selection by $\Delta x_{\ell+1} \in \text{Ker } \varphi'(x_\ell)$. Then, the singular value decomposition of the matrix $\varphi'(x_\ell)$ is made use of again.

Taking account of the freezing zone, one can perform approaching and melting in a unified way. Namely, $\Delta x_{\ell+1}$ is taken randomly in $\text{Ran } \varphi'(x_\ell)^*$ and $\text{Ker } \varphi'(x_\ell)$ according to $\|\Delta x_{\ell+1}\| > \|\varphi'(x_\ell)\| \cdot \mathcal{J}(x_\ell)^{1/2}$ and $\|\Delta x_{\ell+1}\| < \|\varphi'(x_\ell)\| \cdot \mathcal{J}(x_\ell)^{1/2}$, respectively. Here, the singular value decompositions of the matrices $\varphi'(x_\ell)^*$ and $\varphi'(x_\ell)$ are made use of. We call this method the *parallel optimization by random perturbations*.

4 Binding and Streaming

How to lead the melting sequence is mostly important. It should be based on a priori considerations on the expected status. We propose two ways of melting in MEG analysis, that is, *binding* and *streaming*.

In binding, the current elements are supposed to trace the primary current in the form of (9). Therefore, the clustered elements are more desirable, although the number of clustered areas is not prescribed. For this purpose, we adopt the idea of counting measure, which is stated as follows. Namely, given as family of points, we wish to count them. For this purpose, first, we

fix $\varepsilon > 0$ and take the coverings of the set of points by ε balls. Then, the minimal number of those balls is non-decreasing in ε , and converges to the number of points as $\varepsilon \downarrow 0$. In our method, clustered elements in the area with radius ε are regarded as a point. Actual process is stated as follows.

Let $x_\ell = (Q^\ell, a^\ell)$ be in the freezing zone. Here, $Q^\ell = (Q_1^\ell, \dots, Q_N^\ell) \in \mathbf{R}^{3N}$ and $a^\ell = (a_1^\ell, \dots, a_N^\ell) \in \mathbf{R}^{3N}$ denote the multiples of moments and positions of the current elements $\{(Q_k^\ell, a_k^\ell) \mid 1 \leq k \leq N\}$, respectively. Let $S_\ell = \{a_k^\ell \mid 1 \leq k \leq N\}$. First, we take small $\varepsilon > 0$ and provide an ε covering of S_ℓ randomly as $S_\ell \subset \cup_j B(y_j, \varepsilon)$. This process is repeated until a minimal number of covering balls is achieved. We call it *covering*. Next, we reduce each ball as small as possible in such a way as $S_\ell \subset \cup_j B(y_j, \varepsilon_j)$, where $\varepsilon_j \in (0, \varepsilon]$ and $S_\ell^c \cap B(y_j, \delta) \neq \emptyset$ for $\delta \in (0, \varepsilon_j)$. This process is called *biting*. Then, melting for $x_\ell = (Q^\ell, a^\ell)$ is done to get the next status $x_{\ell+1} = (Q^{\ell+1}, a^{\ell+1})$ under the constraint $S_{\ell+1} \subset \cup_j B(y_j, \varepsilon_j)$, where $S_{\ell+1} = \{a_k^{\ell+1} \mid 1 \leq k \leq N\}$ for $a^{\ell+1} = (a_1^{\ell+1}, \dots, a_N^{\ell+1})$. Then, we come back to covering by making $\varepsilon > 0$ smaller. In this method, evaluation of the status of iterative sequences is made by the number of covering balls.

In streaming, one wishes for the distribute current elements to trace the total current density. In this case, it is desirable for the set of current elements to realize a smooth vector field on Ω . Again, we take the covering process to make each current element to recognize the ones near by it.

Namely, taking $\varepsilon > 0$, first we perform the covering process and obtain $S_\ell \subset \cup_j B(y_j, \varepsilon)$. Let S_ℓ^j be the set of positions of elements in $B(y_j, \varepsilon)$, that is, $S_\ell^j = \{a_k^\ell \in S_\ell \mid a_k^\ell \in B(y_j, \varepsilon)\}$. Furthermore, T_ℓ^j and $\tilde{T}_\ell^j$ denote the sets of moments of the elements in $B(y_j, \varepsilon)$ and those coming into $B(y_j, \varepsilon)$, respectively: $T_\ell^j = \{Q_k^\ell \mid a_k^\ell \in S_\ell^j\}$, $\tilde{T}_\ell^j = \{a_k^\ell + Q_k^\ell \in B(y_j, \varepsilon)\}$. Put $\hat{T}_\ell^j = T_\ell^j \cup \tilde{T}_\ell^j$ and

$$d_j = \sum_{Q_k^\ell, Q_{k'}^\ell \in \hat{T}_\ell^j, k \neq k'} \frac{Q_k^\ell \cdot Q_{k'}^\ell}{|Q_k^\ell| |Q_{k'}^\ell|} \quad \text{and} \quad d = \sum_j d_j.$$

Then melting for $x_\ell = (Q^\ell, a^\ell)$ is done to increase d . Now, we get the next step and come back to covering, making $\varepsilon > 0$ smaller. In this method, evaluation of the status of iterative sequences is made by the value d .

Thus, binding and streaming are the processes of resetting the variational structure at each step of iteration. In this sense, they are comparable to the

method of regularization, but the evaluation of the status is more flexible and their ultimately expected forms of $\varepsilon = 0$ are hard to express in variational problems.

5 Further Discussions

In this section, we describe two items. First, the roles of moments and positions are different in approaching, because $E = Q \times \frac{x-a}{|x-a|^3}$ has a singularity at $x = a$. In fact, for $E_j^k(Q, a) = Q_k \times \frac{w_j - a_k}{|w_j - a_k|^3}$ we have

$$E_j^k(Q + \Delta Q, a) = E_j^k(Q, a) + \Delta Q_k \times \frac{w_j - a_k}{|w_j - a_k|^3} \quad (12)$$

and

$$E_j^k(Q, a + \Delta a) = E_j^k(Q, a) - 2Q_k \times \frac{\Delta a_k}{|w_j - a_k|^4} + o(|\Delta a|). \quad (13)$$

This means that the perturbations

$$(w_j - a_k) \times \Delta Q_k \quad \text{and} \quad 2Q_k \times \Delta a_k / |w_j - a_k|$$

are comparable in the contribution of $\Delta_x E_j^k$ for $\Delta \mathcal{J}(x) = \mathcal{J}(x + \Delta x) - \mathcal{J}(x)$. In particular, it happens that

$$|\Delta a_k| \approx \frac{|\Delta Q_k|}{|Q_k|} \cdot |w_j - a_k|^2$$

in $\Delta_x E_j^k$. The contribution of $\Delta_x E_j^k$ to $\Delta \mathcal{J}$ is most important when a_k is very close to w_j by (12) and (13). Therefore,

$$|\Delta a_k| \approx \frac{|\Delta Q_k|}{|Q_k|} \cdot \text{dist}(a_k, \mathcal{O})^2$$

follows to $\Delta \mathcal{J}$, where $\mathcal{O} = \{w_j \mid 1 \leq j \leq M\}$ denotes the set of channels on $\partial\Omega$. If $|\Delta Q_k| \sim |Q_k|$, then $|\Delta a_k| \sim \text{dist}(a_k, \mathcal{O})^2$ and current elements near $\mathcal{O}$ are hard to move. One way to avoid this difficulty is the use of weighted random perturbations. Thus, random perturbations of moments and positions are taken by $1 : \text{dist}(a_k, \mathcal{O})^{-2}$. Another method is to take $|\Delta Q_k| \sim |Q_k| \cdot \text{dist}(a_k, \mathcal{O})^{-2}$. In both cases, one may take $\text{dist}(a_k, \partial\Omega_+)^{-2}$ for $\text{dist}(a_k, \mathcal{O})^{-2}$ if w_j 's are distributed uniformly on the chemi-sphere $\partial\Omega_+$.

The third way is base on the theory of Hooke and Jeeves [9]. Given $k = 1, 2, \dots, N$, we regard φ as a function of $Q_k \in \mathbf{R}^3$. Let $\varphi_k : \mathbf{R}^3 \rightarrow \mathbf{R}^M$ for this mapping. If $Q_k \in \mathbf{R}^3$ attains $\inf_{Q_k \in \mathbf{R}^3} \frac{1}{2} \|\varphi_k(Q_k) - z\|^2$, then it holds that

$$\frac{\partial \varphi_k}{\partial Q_k}^* (Q_k) [\varphi_k(Q_k) - z] = 0$$

similarly to (10). Regarding this, we eliminate $Q \in \mathbf{R}^{3N}$ by

$$\frac{\partial \varphi}{\partial Q_k}^* (Q, a) [\varphi(Q, a) - z] = 0.$$

Because φ_k is affine in Q_k , we can replace $\varphi(Q, a)$ and $\frac{\partial \varphi}{\partial Q_k}^* (Q, a)$ by

$$\varphi(Q_0, a) + \frac{\partial \varphi}{\partial Q_k} (Q_0, a) [Q_k - Q_k^0] \quad \text{and} \quad \frac{\partial \varphi}{\partial Q_k}^* (Q_0, a),$$

respectively, where $Q = (Q_1, \dots, Q_N)$ and $Q_0 = (Q_1^0, \dots, Q_N^0)$ with $Q_\ell = Q_\ell^0$ for $\ell \neq k$. This implies

$$\begin{aligned} \frac{\partial \varphi}{\partial Q_k}^* \cdot \frac{\partial \varphi}{\partial Q_k} (Q_0, a) [Q_k] &= \frac{\partial \varphi}{\partial Q_k}^* (Q_0, a) [z - \varphi(Q_0, a_0)] \\ &\quad + \frac{\partial \varphi}{\partial Q_k}^* (Q_0, a) [Q_k^0]. \end{aligned}$$

Regarding this, we take the following scheme. Namely, given $x_\ell = (Q^\ell, a^\ell)$, we determine the $\ell + 1$ -th moment $Q^{\ell+1} = (Q_1^{\ell+1}, \dots, Q_N^{\ell+1})$ as a function of $a^{\ell+1} \in \mathbf{R}^N$ by

$$\begin{aligned} \frac{\partial \varphi}{\partial Q_k}^* \cdot \frac{\partial \varphi}{\partial Q_k} (Q^\ell, a^\ell) [Q_k^{\ell+1}] &= \frac{\partial \varphi}{\partial Q_k}^* (Q^\ell, a^\ell) [z - \varphi(Q^\ell, a^\ell)] \\ &\quad + \frac{\partial \varphi}{\partial Q_k}^* (Q^\ell, a^\ell) [Q_k^\ell] \end{aligned}$$

for $k = 1, 2, \dots, N$. Then, one of the optimization processes described in the previous section is applied for $\mathcal{J}$ with Q eliminated in this way. The linear part in Q_k of $\varphi_k(Q_k)$ is given by a triple product and 3×3 matrix $\frac{\partial \varphi}{\partial Q_k}^* \frac{\partial \varphi}{\partial Q_k}$ is easy to calculate. This method is applicable also to the over-determined case.

Turning to the next item, we see that $\Delta_Q E_j^k \cdot \nu_j = 0$ and $\Delta_a E_j^k \cdot \nu_j = o(1)$ if Q_k and $w_j - a_k$ are parallel to ν_j , respectively. We say that an element

(Q_k, a_k) is in the *degenerate hole* if both conditions are satisfied for very close $w_j \in \mathcal{O}$. It is easy to see that such an element is hard to move in approaching. On the other hand, $|\Delta_Q E_j^k \cdot \nu_j|$ and $|\Delta_a E_j^k \cdot \nu_j|$ become maximum if ΔQ_k and Δa_k are perpendicular to ν_j , respectively. We say that an element (Q_k, a_k) is in the *looping hole* if ν_j , $a_k - w_j$, and Q_k are perpendicular to very close $w_j \in \mathcal{O}$. In this case the contribution from ΔE_j^k to reduce $\mathcal{J}$ becomes maximum when those conditions are preserved. This situation can keep to hold. Roughly speaking, those holes make $\mathcal{J}$ to be flat outside the freezing zone. Pattern formation is observed sometimes in the transient states of the method of current element distribution, and holes can be its reason.

6 Numerical Results

Numerical experiments for the current element distribution method (without melting) have been tried for several years. In [12], [14], [1], it is reported that circular like total current densities are well traced by this method.

The melting was first proposed in [16]. There, a more rough method was adopted, and only one ball is taken to cover $\mathcal{S}_\ell$. Based on this covering, biting and melting are done similarly. It was seen that current elements are clustering near the originally clustered dipoles.

Figure 1 of the present paper illustrates the efficiency of employing such a step size control for Δa_k and ΔQ_k as $|\Delta Q_k| \sim |Q_k| \cdot \text{dist}(a_k, \mathcal{O})^{-2}$. It is seen that the accuracy is improved rapidly.

Figures 2, 3, and 4 show the effect of binding. There, it is seen that the current elements are clustering to three areas where the original dipoles are set.

Further programs are the following.

1. Examining [12], [14], [1]. Are they freezing ?
2. The uses of parallel optimization, binding, and streaming.
3. Trying several treatments for moments.

Part of this study is funded by the Research for the Future (RFTF) Program of Japan Society for the Promotion of Science, Integrated Fields Area: Research on Biomedical Observation and Control, "Application of Heuristic Solution of Inverse Problem to Medical

Diagnoses and Biological Research". The second author thanks Dr. M. Higuchi for pointing out several references and Professor G. Uehara for some comments on the descriptions on SQUID.

References

- [1] Y. ADACHI, M. SHIMOGAWARA, Y. HARUTA, M. HIGUCHI, and H. KADO, *A current element distribution method applied to magnetoencephalography*, Abstract for Intermag 99.
- [2] A. DALE and M. SERENO, *Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach*, J. Cogn. Neurosci. **5** (2), 162–176 (1993)
- [3] A.S. FOKAS, I.M. GEL'FAND, and Y. KURYLEV, *Inversion method for magnetoencephalography*, Inverse Problems **12** (1996) L9-L11.
- [4] A.S. FOKAS and Y. KURYLEV, *On the determination of neuronal currents via magnetoencephalography*, preprint
- [5] D.B. GESELOWITZ, *On a magnetic field generated outside an inhomogeneous volume conductivity by internal current sources*, IEEE Trans. Magn. **MAG-6**, 346-347 (1970)
- [6] D.B. GESELOWITZ and F. GRZYNSZPAN, *Model studies of the magnetocardiogram*, Biophys. J. **13**, 911-925 (1973)
- [7] , M.S. HÄMÄINEN and R.J. ILMONIEMI, *Interpreting magnetic fields of the brain: minimum norm estimates*, Med. Biol. Eng. Comput. **32**, 35-42 (1994)
- [8] M.S. HÄMÄINEN, R.J. ILMONIEMI, and J. SARVAS, *Interdependence of information conveyed by the magnetoencephalogram and the electroencephalogram*, In; Harrio, H. (ed.), *Theory and Applications of Inverse Problems*, Research Notes in Math. **167**, Pitman, 1988, pp. 27-37.
- [9] R. HOOKE and T.A. JEEVES, *Direct search solution of numerical and statistical problem*, J. Assoc. Comput. Mach. 212-229, (1961)

- [10] M. HUANG, et. al., *Multi-start downhill simplex method for spatio-temporal source localization in magnetoencephalography*, *Electroenceph. Clin. Neurophysiol.* **108**, 32-44 (1998).
- [11] R.I. ILMONIEMI, *Magnetoencephalography - a tool for studies of information processing in the human brain*, In; *The Inverse Problem* (H. Lübbig ed.), Akademie, 1995.
- [12] H. KADO, M. KASHIWAYA, M. HIGUCHI, and H. MIMURA, *Direct approach to an inverse problem: a trial to describe signal sources by current elements distribution*, *Advances in Biomagnetism*, Plenum Press, New York, pp. 579-582 (1989).
- [13] J. SARVAS, *Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem*, *Phys. Med. Biol.* **32**, 11-22 (1987)
- [14] M. SHIMOAWARA, H. KADO, H. KOHNO, and M. HIGUCHI, *Magnetic source imaging by current element distribution*, *Biomagnetism: Clinical Aspects*, Elsevier, New York, pp. 757-760 (1992).
- [15] J. STOER and R. BULIRSCH,, *Introduction to Numerical Analysis*, 2nd edition, Springer, New York, 1993.
- [16] T. SUZUKI, Y. ADACHI, H. KADO, N. SAITO, and M. SHIMOAWARA, *Mathematical foundations for some numerical methods in magnetoencephalography*, In; J. Nenonen, R.J. Ilmoniemi, and T. Katila, eds. BIOMAG2000, Proc. 12th Int. Conf. on Biomagnetism, pp. 0644-0646, Helsinki Univ. of Technology, Espo, Finland, 2001.
- [17] K. UUTELA, et. al., *Global optimization in the localization of neuromagnetic sources*, *IEEE Trans. Biomed. Eng.* **45**, 716-723 (1998).

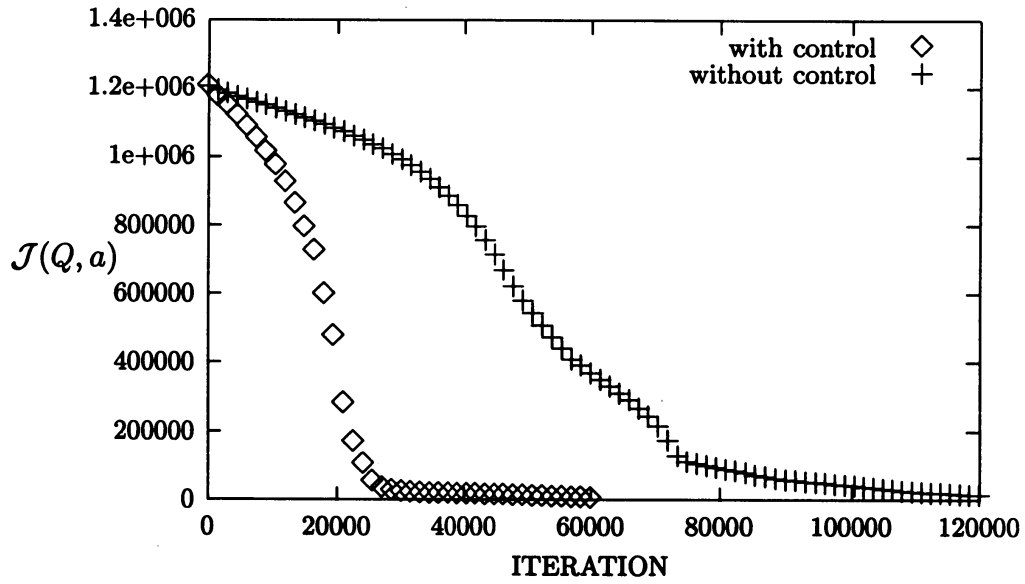


Figure 1: Approaching process with a step size control $|\Delta Q_k| \sim |Q_k| \cdot \text{dist}(a_k, \mathcal{O})^{-2}$ ($\diamond$) and without it ($+$). The control reduces the number of iterations by quarter.



Figure 2: Initial state. 20 channels and 80 current elements are randomly distributed.

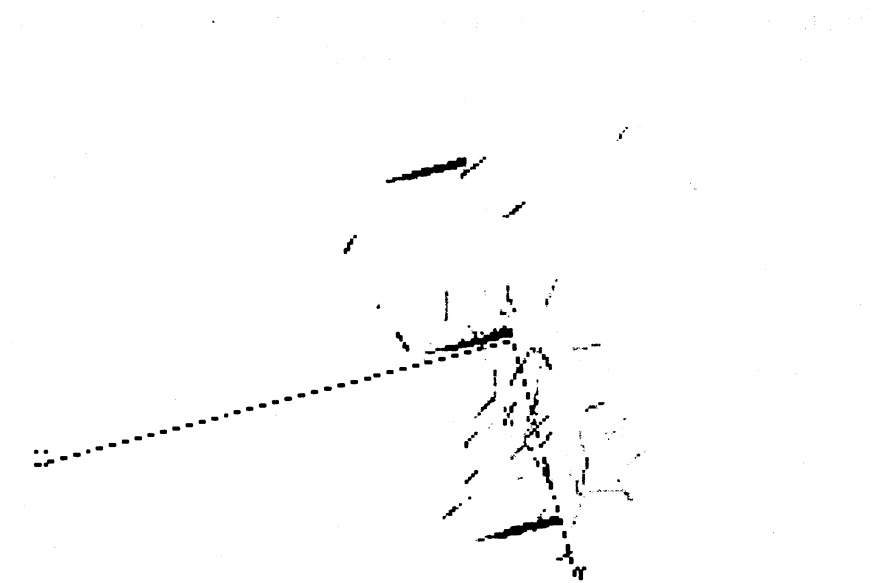


Figure 3: State after approaching and melting process.



Figure 4: State after approaching and melting process. (from another viewpoint).