

Linear Inverse Solutions With Optimal Resolution Kernels Applied to Electromagnetic Tomography

Rolando Grave de Peralta Menendez,^{1*} Olaf Hauk,² Sara Gonzalez Andino,¹ Hermann Vogt,² and Christoph Michel¹

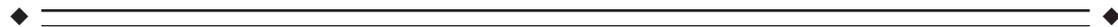
¹Functional Brain Mapping Laboratory, University Hospital Geneva, 1211 Geneva, Switzerland

²Department of Clinical Neurophysiology, University of Magdeburg, Magdeburg 39120, Germany



Abstract: This paper discusses the construction of inverse solutions with optimal resolution kernels and applications of them in the reconstruction of the generators of the EEG/MEG. On the basis of the framework proposed by Backus and Gilbert [1967], we show how a family of well-known solutions ranging from the minimum norm method to the generalized Wiener estimator can be derived. It is shown that these solutions have optimal properties in some well-defined sense since they are obtained by optimizing either the resolution kernels and/or the variances of the estimates. New proposals for the optimization of resolution are made. In particular, a method termed “weighted resolution optimization” (WROP) is introduced that deals with the difficulties inherent to the method of Backus and Gilbert [1967], from both a conceptual and a numerical point of view. One-dimensional simulations are presented to illustrate the concept and the interpretation of resolution kernels. Three-dimensional simulations shed light on the resolution properties of some linear inverse solutions when applied to the biomagnetic inverse problem. The simulations suggest that a reliable three-dimensional electromagnetic tomography based on linear inverse solutions cannot be constructed, unless significant a priori information is included. The relationship between the resolution kernels and a definition of spatial resolution is emphasized. Special consideration is given to the use of resolution kernels to assess the properties of linear inverse solutions as well as for the design of inverse solutions with optimal resolution kernels. *Hum. Brain Mapping* 5:454–467, 1997. © 1997 Wiley-Liss, Inc.

Key words: inverse problem; electroencephalography; magnetoencephalography; resolution kernels; spatial resolution; electromagnetic tomography



INTRODUCTION

Although many complementary noninvasive imaging techniques for the study of the human brain have been developed in the last few decades, the only

modality that provides the temporal resolution needed to study the dynamics of brain processes is the measurements of the electric or magnetic fields produced by neuronal currents on or over the scalp surface. The major limitation of these techniques appears when we attempt to construct a three-dimensional tomography of these currents, which leads to an intrinsically non-unique inverse problem. It is well-known that neither an infinite number of exact electric or magnetic measurements alone, nor a combination of them, determine uniquely the current distribution within the brain [Hämäläinen et al., 1993]. Additionally, the data are in practice noisy and insufficient, which results in a

Contract grant sponsor: Swiss National Foundation; Contract grant numbers: 4038-044081/1, 32-37819.93; Contract grant sponsor: Deutsche Forschungsgemeinschaft.

*Correspondence to: Rolando Grave de Peralta Menendez, Functional Brain Mapping Laboratory, University Hospital Geneva, 24, Rue Micheli du Crest, 1211 Geneva 4, Switzerland. E-mail: grave@diogenes.hcuge.ch

Received for publication 9 June 1997; accepted 21 July 1997

reduction of the reliability of the solution: the former introduces statistical uncertainties in the model, while the latter smoothes out the spatial details. Even so, the analysis of reliability is relevant for any inversion method; it is relatively easy to define measures to quantify the properties of linear inversion methods concerning resolution and stability, and to design estimators based on the optimization of these measures. While the methods to deal with the instability are well-developed [Bertero et al., 1988], resolution seems to deserve further attention.

Even if different interpretations of resolution exist, we will refer to resolution as the capability to resolve (detect) details on a certain length scale. It is intuitively clear that the measurements reflect some properties of the brain sources, but the important question is: "Which features of brain activity can be identified from the data without further knowledge about the sources?" Though there is no definite answer, we can be sure that there is a resolution limit that can not be surpassed by any method. While the measurements might allow us to determine changes at the scale of the whole brain or of one hemisphere, changes at neuronal scales will certainly not be detected. In general, for any point in the source space, the resolution is different, depending on the noise, the sensor configuration, and the electromagnetic properties of the medium (lead fields), e.g., silent sources cannot be retrieved at all, whereas (nonsilent) sources near to the sensors are expected to be better resolved.

Linear and nonlinear inversion methods have been in use for some time now to yield a three-dimensional estimation of the current distribution within the human brain [Clarke et al., 1989; Ioannides et al., 1989; Pascual-Marqui et al., 1995]. Different methods have been explored [reviewed in Hämäläinen et al., 1993; George et al., 1995], but until recently, these methods were introduced mainly to construct sources with some predefined global properties (minimum norm, smoothness). The fact that linear inversion methods yield estimates of the source activity that can be interpreted as weighted averages has rarely been reflected in the literature [Robinson and Rose, 1992; Dale and Sereno, 1993]. Recently, we have used this interpretation for several purposes: 1) Analysis of capabilities and limitations of linear inverse solutions or comparisons among them [Grave de Peralta Menendez et al., 1996; Hauk et al., 1996; Lütkenhöner and Grave de Peralta Menendez, 1997] and 2) Proposal of new inverse solutions based on the estimation of averages [Grave de Peralta Menendez and Gonzalez Andino, 1997].

For the case of linear inversion methods, the achievable resolution can be analyzed by means of the

resolution or averaging kernels [Backus and Gilbert, 1967, 1968, 1970; Grave de Peralta Menendez et al., 1996; Grave de Peralta Menendez and Gonzalez Andino, 1997], which, associated with every linear estimator, describe how all possible active sources influence the estimate. If an estimate of the source activity at a certain location is desired, the influence of sources distant from this location should be small compared to closer ones. In this case the resolution kernels peak around the target point and the width of this peak is directly related to the resolution at this point. One goal of this paper is to show how the quality of a resolution kernel and the stability of the estimates can be quantified in order to derive optimal inverse solutions. Since there is no unique way to measure the goodness of a resolution kernel in terms of its closeness to the ideal one, different measures of the optimality of the resolution kernel will lead to different solutions.

In the first section of the paper, the basic theory of the bioelectromagnetic inverse problem is briefly examined. Also, the concept of the model resolution matrix and a description of the general methodology that can be used to obtain solutions with optimal resolution kernels is introduced. With these elements, some linear inverse solutions are cast in this framework and two new solutions, the weighted resolution optimization (WROP) method and a generalization of the method of Capon [1969], are proposed. The case of noise is considered, and it is shown how inclusion of the variance of the solution in the optimization of linear inversion methods leads to a trade-off between stability and resolution. Computer simulations are used to facilitate the comprehension of some concepts presented in the theoretical section, and to illustrate the difficulties inherent to the bioelectromagnetic inverse problem in an approximate model of the head. The analysis of solutions is based on their resolution kernels, which are independent of the data. Thus, the conclusions derived in this section will not change by considering the noise. The last section is devoted to the discussion of the results, considering their possible applications, alternatives, and future trends.

In this paper, vectors will be represented by lower-case bold letters, and matrices by uppercase. All vectors are column vectors; \mathbf{W}_{ij} represents the element in row i and column j of \mathbf{W} . \mathbf{W}_i and \mathbf{W}_j stand for the column vectors determined by the i -th row and the j -th column of \mathbf{W} , respectively, and $\text{diag}(\mathbf{W})$ stands for the diagonal matrix with the same elements as in the main diagonal of \mathbf{W} . If \mathbf{x} is a vector, then $\tilde{\mathbf{x}}$ represents the diagonal matrix with elements according to \mathbf{x} .

PROBLEM STATEMENT

The relationship between electric potentials or magnetic fields measured at/near the scalp surface and their sources can be expressed in terms of a linear integral equation [Sarvas, 1987]:

$$\mathbf{d} = \int \mathbf{L}(\vec{r})\vec{j}(\vec{r})d\vec{r} + \mathbf{n} \quad (1)$$

where \mathbf{d} is the m -dimensional vector of measured data points and $\vec{j}(\vec{r})$ is the source current density. The number of components n_c of $\vec{j}(\vec{r})$ depends upon the constraints imposed on the currents in the selected source space. In the most general case n_c is equal to 3 but if, for example, the source space is selected as a two-dimensional surface approximating the cortical sheet and the currents are constrained to be normal to this surface, then n_c is equal to 1. The m by n_c lead field matrix $\mathbf{L}(\vec{r})$ reflects the sensitivity of every sensor to the sources, and \mathbf{n} stands for the noise which is assumed to be additive and with zero mean. The integral is taken over the whole source space. In practice, the problem is usually discretized [Greenblatt, 1993], e.g., the source current is approximated by a large number of current dipoles of arbitrary orientation at known locations. This results in the linear matrix equation:

$$\mathbf{d} = \mathbf{L}\mathbf{j} + \mathbf{n} \quad (2)$$

where \mathbf{d} is again the m -dimensional vector of measured data points, \mathbf{j} represents the discretized source current density consisting of p unknowns, and \mathbf{n} stands for the noise as above. Each column of the discrete lead field matrix \mathbf{L} contains the forward solution for one component of a specific dipole. Therefore, every measurement \mathbf{d}_i in a specific channel i can be interpreted as a projection of the source \mathbf{j} on to the i -th row of the lead field matrix plus a random noise term n_i . For simplicity, we will here consider the discrete case. Nevertheless, the solutions discussed in this paper are not restricted to the points used in the discretization. This aspect will be further explained in the discussion.

In the discrete inverse problem we attempt to find \mathbf{j} from the knowledge of the data \mathbf{d} and the lead field matrix \mathbf{L} . Since in realistic situations $m \ll p$, this is a highly underdetermined inverse problem. The data contain information only about the m -dimensional subspace, whereas the source space is p -dimensional (infinite in the continuous case). This problem is comparable to the reconstruction of a three-dimensional object from its shadow, which is a two-

dimensional projection of the object. Nevertheless, even in highly underdetermined problems, the data reflect some properties of the sources. The cornerstone to determine the inferences that can be drawn from the data by means of linear inverse solutions is the concept of the model resolution matrix.

The resolution matrix

Any linear inverse solution to Equation (2) can be written as

$$\hat{\mathbf{j}} = \mathbf{G}\mathbf{d} \quad (3)$$

Substitution of the data \mathbf{d} according to Equation (2) into Equation (3) leads to a fundamental equation for underdetermined linear systems:

$$\hat{\mathbf{j}} = \mathbf{G}\mathbf{d} = \mathbf{G}\mathbf{L}\mathbf{j} + \mathbf{G}\mathbf{n} = \mathbf{R}\mathbf{j} + \mathbf{e} \quad (4)$$

where $\mathbf{R} = \mathbf{G}\mathbf{L}$ is the resolution matrix [Menke, 1989; Grave de Peralta Menendez et al., 1996] and $\mathbf{e} = \mathbf{G}\mathbf{n}$ stands for the part of the solution due to noise in the data. If we consider, for simplicity, the case without noise, i.e., $\hat{\mathbf{j}} = \mathbf{R}\mathbf{j}$, an important relationship between the estimate $\hat{\mathbf{j}}$ and the real source \mathbf{j} can be seen. Every component of $\hat{\mathbf{j}}$ is a weighted average of the actual \mathbf{j} , with the weighting factors given by the elements of the rows of the resolution matrix \mathbf{R} . For this reason, the rows of \mathbf{R} are called averaging kernels. Unique, perfectly accurate estimates of the actual sources are obtained when $\mathbf{R} = \mathbf{I}$ (the identity matrix). However, this does not occur in underdetermined inverse problems since the matrix \mathbf{R} cannot have a rank greater than m (the number of sensors). The best resolution matrices we could hopefully obtain are matrices with rows peaked around the point r_0 , for which they give an estimate and small amplitudes elsewhere. In such cases, the estimates can be interpreted as smoothed versions of the actual current distribution, and the degree of smoothing depends on the shape of the averaging kernels. Note that in this interpretation in terms of averages, it is implicitly assumed that simultaneously active sources at remote points have small influence in the estimate of the source under analysis. In this case where the averaging kernel has a peak around r_0 , the estimate at this point will be an average of the real source around r_0 . Backus and Gilbert [1967] identified the width of the peak with the concept of resolution, i.e., a rough measure of the finest details (highest spatial frequencies) which can be resolved by the data around r_0 . For that reason the rows of \mathbf{R} are

also called resolution or resolving kernels. In practice, the resolution kernels may exhibit large sidelobes (see Computer Simulations, below) or are incorrectly peaked at locations distant from the target point r_0 . All these aspects confuse the interpretation of the resolution concept.

Equipped with this appealing interpretation, Backus and Gilbert [1968] concluded that optimal resolution kernels should have the sharpest possible peak around the target point r_0 , compatible with good suppression of contributions from distant r . In other words, they should be close to the ideal one: the delta function for the continuous case, or the rows of the identity matrix in the discrete case. Backus and Gilbert [1968] defined a numerical measure for the closeness to the delta function, termed the spread, which, when optimized, leads to solutions with optimal resolution kernels. The resultant optimization problem, and a family of solutions that can be derived by considering the discrete spread of Backus and Gilbert [1968], are the topics of the following section.

Optimization of the resolution matrix

We will consider the discrete Backus and Gilbert spread [Menke, 1989] to measure the distance to the ideal resolution matrix, i.e., the identity matrix \mathbf{I} :

$$s(\mathbf{R}, \mathbf{I}) = \sum_i \sum_j W_{ij} (\mathbf{R}_{ij} - \mathbf{I}_{ij})^2 \quad (5)$$

where the W_{ij} are positive otherwise arbitrary weighting factors.

After some algebraic transformations and considering that $\mathbf{R} = \mathbf{GL}$, Equation (5) can be rewritten as:

$$\begin{aligned} s(\mathbf{GL}, \mathbf{I}) &= \sum_i (\mathbf{G}_i^t \mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}^t \mathbf{G}_i - 2\mathbf{W}_{ii} \mathbf{G}_i^t \mathbf{L}_i + \mathbf{W}_{ii}) \\ &= \sum_i (\mathbf{G}_i^t \mathbf{S}_i \mathbf{G}_i - 2\mathbf{W}_{ii} \mathbf{G}_i^t \mathbf{L}_i + \mathbf{W}_{ii}) \end{aligned} \quad (6)$$

where $\mathbf{S}_i = \mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}^t$ is called the spread matrix. The summation index runs over the resolution kernels, i.e., the rows of the resolution matrix. Note that the i -th term in the summation represents the distance from the i -th resolution kernel to the ideal one and depends only upon the i -th row of the inverse matrix \mathbf{G} . This property facilitates the computation of the solutions derived in what follows, since they are determined by optimizing the spread $s(\mathbf{GL}, \mathbf{I})$ with respect to the elements of \mathbf{G} , using additional constraints when needed to ensure unique estimators.

Minimum norm method

If all the weights are equal to a positive arbitrary value, i.e., $W_{ij} = \alpha > 0$, the spread is just a multiple of the Frobenius norm. For any row we have the following optimization problem:

$$\min \mathbf{G}_i^t \mathbf{S}_i \mathbf{G}_i - 2\alpha \mathbf{G}_i^t \mathbf{L}_i + \alpha \text{ with solution} \quad (7)$$

$$\mathbf{G}_i = \alpha \mathbf{S}_i^{-1} \mathbf{L}_i = (\mathbf{L} \mathbf{L}^t)^{-1} \mathbf{L}_i \quad (8)$$

which corresponds to the rows of the minimum norm (MN) inverse matrix or Moore-Penrose pseudoinverse [Penrose, 1955], applied in the bioelectromagnetic inverse problem by different groups [e.g., Hämäläinen and Ilmoniemi, 1984; Wang et al., 1992].

Backus and Gilbert method

Since Robinson and Rose [1992] mentioned the theory of Backus and Gilbert [1967], we repeatedly suggested different applications of it to bioelectromagnetic data. In Grave de Peralta Menendez et al. [1996] and Grave de Peralta Menendez and Gonzalez [1997], we introduced the model resolution matrix as the basis for the analysis, comparison, and design of linear inverse solutions. Other applications are described in Hauk et al. [1996], Lütkenhöner and Grave de Peralta Menendez [1997], and Hauk [1996]. The key idea of the method of Backus and Gilbert [1968], to obtain smooth and compact resolution kernels, is a selection of weights that increasingly penalize the amplitudes of the kernels at points distant from the one under analysis. For the sake of clarity, we will consider, below, their proposals of weights for scalar and vector fields separately.

Scalar fields

For the case of scalar fields the weights are selected such that $\mathbf{W}_{ij} = f(d_{ij})$, where f is a monotone function such that $f(0) = 0$ and d_{ij} is the distance between the solution points associated with the unknowns i and j . In this way, the resolution kernel for the unknown i is compelled to have lower amplitudes at distant points and larger amplitudes near the point under estimation. Probably, the most widely used weighting function is $f(x) = x^2$, which leads to the so-called second moment norm [Treitel and Lines, 1982].

Since $\mathbf{W}_{ii} = f(0) = 0$, the optimization of the spread is not enough to uniquely determine \mathbf{G}_i . Thus an additional restriction is required. Backus and Gilbert [1968] con-

strained the integral of a resolution kernel to be equal to one (unimodularity), which reduces to the condition $\mathbf{R}_i^t \vec{\mathbf{1}} = \mathbf{G}_i^t \mathbf{L}_i \vec{\mathbf{1}} = 1$ in the discrete case, where $\vec{\mathbf{1}}$ is a vector containing only ones as elements. The resultant optimization problem is given by:

$$\begin{aligned} \min \mathbf{G}_i^t \mathbf{S}_i \mathbf{G}_i \\ \text{s.t. } \mathbf{G}_i^t \mathbf{u} = 1 \end{aligned} \quad (9)$$

with $\mathbf{u} = \mathbf{L}_i \vec{\mathbf{1}}$, i.e., the k -th component of \mathbf{u} is the sum over the elements of the k -th row of \mathbf{L} .

This leads to the solution:

$$\mathbf{G}_i = \frac{\mathbf{S}_i^{-1} \mathbf{u}}{\mathbf{u}^t \mathbf{S}_i^{-1} \mathbf{u}} = \frac{(\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{L}_i \vec{\mathbf{1}}}{\vec{\mathbf{1}}^t \mathbf{L}^t (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{L}_i \vec{\mathbf{1}}} \quad (10)$$

A more general formulation of the constraint is given by selecting the vector and the constant arbitrarily for each row, i.e., $\mathbf{G}_i^t \mathbf{u}_i = \alpha_i$. Then the solution generalizes to:

$$\mathbf{G}_i = \alpha_i \frac{\mathbf{S}_i^{-1} \mathbf{u}_i}{\mathbf{u}_i^t \mathbf{S}_i^{-1} \mathbf{u}_i} = \alpha_i \frac{(\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{u}_i}{\mathbf{u}_i^t (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{u}_i} \quad (11)$$

An interesting alternative to the unimodularity constraint is to impose that the diagonal element of the row of the resolution matrix is equal to 1. With this restriction we attempt to determine the actual strength of a source at that location if no others are simultaneously active. This constraint can be expressed by $\mathbf{u}_i = \mathbf{L}_{,i}$ and $\alpha_i = 1$, which leads to:

$$\mathbf{G}_i = \frac{\mathbf{S}_i^{-1} \mathbf{u}_i}{\mathbf{u}_i^t \mathbf{S}_i^{-1} \mathbf{u}_i} = \frac{(\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{L}_{,i}}{\mathbf{L}_{,i}^t (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L})^{-1} \mathbf{L}_{,i}} \quad (12)$$

Vector fields

For a vector field of dimension N , Backus and Gilbert [1968] suggested using weights

$$\begin{aligned} W_{ij} = f(d_{ij}) \delta_{rs} + (1 - \delta_{rs}), \quad \text{where } i = (p - 1)N + r, \\ j = (q - 1)N + s, \quad 1 \leq r, s \leq N \end{aligned}$$

represent the r -th and s -th components on points p and q , respectively, and δ_{rs} stands for the discrete Dirac delta function, with value one for $r = s$ and zero otherwise.

This proposal of weights for the vector case is not in full agreement with the original idea of the Backus and

Gilbert method to derive compact resolution kernels. For example, consider the estimation of the r -th component of a source at the p -th point. For any point q , the two components that differ from the one under estimation ($s \neq r$), have constant weights one, independently of the Euclidean distance between p and q . The only component that has weights according to this distance is the r -th itself. However, according to the original philosophy of Backus and Gilbert, the influence of any source component at remote points should be penalized. Thus, we suggest weights that depend on the Euclidean distances for all three components, and adding an arbitrary constant to the ‘‘other’’ components ($s \neq r$) to penalize even more their influence, i.e.,

$$W_{ij} = f(d_{ij}) + (1 - \delta_{rs})\alpha$$

with $\alpha > 0$. We will refer to these weights as corrected weights and the solution obtained using this weighting strategy as corrected Backus and Gilbert (CB&G). For the case of scalar fields, CB&G is the same as the original Backus and Gilbert (B&G).

Weighted resolution optimization (WROP) method

The unimodularity condition selected by Backus and Gilbert [1968] to guarantee unique estimators is only one of the possible constraints fulfilled by an ideal resolution matrix. In fact, the particular properties of the problem under consideration could suggest different desirable attributes for the resolution kernels. Since it is generally sufficient to select weights that determine the value of the diagonal element \mathbf{W}_{ii} to obtain unique estimates, we propose the use of more flexible weights. In particular, for the case of scalar fields we propose the use of $\mathbf{W}_{ij} = f(d_{ij}) + \beta$, $\beta > 0$, with f as in the Backus and Gilbert method. With this selection $\mathbf{W}_{ii} = \beta$, and \mathbf{G}_i can be determined solving the following optimization problem without additional constraints:

$$\min \mathbf{G}_i^t \mathbf{S}_i \mathbf{G}_i - 2\mathbf{W}_{ii} \mathbf{G}_i^t \mathbf{L}_{,i} + \mathbf{W}_{ii} \quad (13)$$

with solution

$$\mathbf{G}_i = \mathbf{W}_{ii} \mathbf{S}_i^{-1} \mathbf{L}_{,i} = \mathbf{W}_{ii} (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}^t)^{-1} \mathbf{L}_{,i} \quad (14)$$

For the estimation of vector fields the weights are selected on two bases: to eliminate the need for additional constraints and to diminish the inconsistencies of Backus and Gilbert’s original selection de-

scribed in the previous section, i.e.,

$$W_{ij} = [f(d_{ij}) + \beta] + (1 - \delta_{rs})\alpha$$

Here α is a positive constant, and the remaining indices are as in the previous section.

In both cases, for scalar and vector fields, the spread function for this method can be decomposed in two terms, namely a combination of the minimum norm spread \mathbf{S}_i^{MN} and the corrected Backus and Gilbert spread $\mathbf{S}_i^{\text{CB\&G}}$

$$\begin{aligned} s(\mathbf{G}\mathbf{L}, \mathbf{I}) &= \sum_i (\mathbf{G}_i^t \mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}^t \mathbf{G}_i - 2\beta \mathbf{G}_i^t \mathbf{L}_i + \beta) \\ &= \sum_i (\mathbf{S}_i^{\text{MN}} + \lambda \mathbf{S}_i^{\text{CB\&G}}) \end{aligned} \quad (15)$$

where the coefficient $\lambda = 1/\beta$ represents the regularization parameter, such that the solution tends to the minimum norm solution when λ tends to zero [Bertero et al., 1988]. The addition of the second term will in general improve the compactness and smoothness of the resolution kernels obtained by the minimum norm method (see Computer Simulations, below).

Note that Equation (14) corresponds to the rows of a weighted minimum norm and when the parameter β depends on the index j , the WROP method corresponds to a regularization with respect to a weighted minimum norm solution. Variations of β with both indexes (i, j) can also be considered, illustrating the flexibility of the WROP method.

Generalized Capon method

The method proposed by Capon [1969] (see also Sekihara and Scholz [1996]) is an alternative formulation to find optimal estimators. In this method one attempts to find a vector \mathbf{G}_i , which “filters out” the activity that arises from one special location, and suppresses the activity from all others. It should be parallel to the forward solution of the target dipole at the location of interest and at the same time orthogonal to the forward solution of all the other dipoles. This idea can be modified according to the theory of Backus and Gilbert [1967], i.e., weighting the forward solutions concerning their Euclidean distances to the location of interest. This can be stated as:

$$\max \frac{(\mathbf{G}_i^t \mathbf{L}_i)^2}{\sum_k \mathbf{W}_{ik} (\mathbf{G}_i^t \mathbf{L}_k)^2} = \max \frac{\mathbf{G}_i^t \mathbf{L}_i \mathbf{L}_i^t \mathbf{G}_i}{\mathbf{G}_i^t (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}) \mathbf{G}_i} \quad (16)$$

with solution

$$\mathbf{G}_i = \alpha (\mathbf{L} \tilde{\mathbf{W}}_i \mathbf{L}^t)^{-1} \mathbf{L}_i = \alpha \mathbf{S}_i^{-1} \mathbf{L}_i \quad (17)$$

Since a ratio is maximized, the result is still undetermined in a scaling factor α . This, however, does not influence the shape of the resolution kernels. Though the expression to be maximized appears to be very different from the ones presented above (it is even nonlinear in \mathbf{G}_i), its solution is proportional to the Backus and Gilbert solution with the constraint that the diagonal elements of the resolution matrix are equal to one, i.e., it is equal to Equation (12) up to a scaling factor.

Regularization and/or noisy data

So far we have mainly described possibilities to minimize the influence of simultaneously active sources on the estimators, especially those at remote sites. However, real measurements are contaminated with signals that do not arise from the brain sources but from the sensors, the recording system, etc. These signals also influence the inverse solution and may even be amplified, which could produce severely distorted or unstable estimates of real activity. The behavior of an inverse solution in the presence of noise is referred to as stability and is closely related to the numerical ill-conditioning of the spread matrix for the solutions described here. The most commonly used procedure to increase the stability of the solution is to add a “regularizer” matrix \mathbf{M} to the spread matrix \mathbf{S}_i , or equivalently to minimize a combination of the spread in Equation (6) and a quadratic form associated with variance of the estimates [Backus and Gilbert, 1970], given by:

$$\mathbf{G}_i^t \mathbf{M} \mathbf{G}_i \quad (18)$$

Since it is usually impossible to minimize both terms, the spread in Equation (6) and the variance of the estimates in Equation (18) at once, a compromise between the minimization of the spread of the resolution kernels and the variance of the estimators has to be made. The relative weight given to each term in this trade-off is controlled by the so-called regularization parameter λ , $0 \leq \lambda \leq 1$, whose optimum value can be determined using the trade-off curve of Backus and Gilbert [1970], or by standard methods [Bertero et al., 1988]. Note that when this quadratic form is added to the minimization of the spread, the optimization problems and the solutions described so far (Equations

7–17), remain the same, but S_i is substituted by $(1 - \lambda)S_i + \lambda M$.

Since the regularizer matrix M is introduced to cope with the noise, M should include as much information on its statistical properties as possible. If the covariance matrix of the noise Σ is known, then according to Equation (4), the contribution of the noise can be estimated by:

$$e_i^2 = \|G_i^t n\|^2 \cong G_i^t \Sigma G_i \text{ and then we select } M = \Sigma. \quad (19)$$

If only the variance of the measurement is known, we can use $M = \text{diag}(\Sigma)$ and, in general, using an a priori selected regularizer C we can define $M = C$. If no information is available, the choice $M = I$, which corresponds to the Tikhonov-Phillips regularization, is appropriate.

For illustration purposes, let's assume that the covariance matrix of the noise is known and all the weight is given to the minimization of the influence of noise, i.e., $\lambda = 1$, $M = \Sigma$. If, additionally, we constrain the diagonal element of the resolution matrix to be one (see Equation 12), we obtain the following optimization problem:

$$\begin{aligned} \min G_i^t \Sigma G_i \\ \text{s.t. } G_i^t L_i = 1 \end{aligned} \quad \text{with solution } G_i = \frac{\Sigma^{-1} L_i}{L_i^t \Sigma^{-1} L_i} \quad (20)$$

This solution coincides with the generalized Wiener estimator derived by Sekihara and Scholz [1996] under the hypothesis that the generators are uncorrelated. This solution is not meant to optimize the resolution kernels, but it optimally separates the forward solution of one dipole from the noise in the measured data.

COMPUTER SIMULATIONS

The computer simulations described in this section have two main goals: 1) To illustrate the concept and utility of the resolution kernels using an oversimplified one-dimensional (1D) model, and 2) to analyze the possibilities of linear inversion methods in terms of its resolution kernels. This division into 1D and three-dimensional (3D) simulations is mainly due to the difficulties in graphically representing a resolution kernel for a vector field (the current density j) in a three-dimensional grid of solution points. Accordingly, the initial 1D simulations consider a planar model for the sensors and solution points and a monopolar model for the sources. Selecting the sources as dipoles poses the problem of representing a vector field in two dimensions which obscures the interpretation of the

resolution kernels, the basic goal of this section. The interpretation of the resolution kernels and how they can be used to predict the properties of a linear solution are the basis of understanding the more realistic 3D simulations. In the subsection entitled Three-Dimensional Simulations, it is shown how the Backus and Gilbert resolution kernels are improved by using a weighting strategy as in the WROP method. Also, it is shown that there are brain sites where poor-resolution kernels are obtained even with solutions that explicitly optimize resolution.

The independence of the resolution matrix from the data, i.e., the independence of the resolution kernels (see Equation 3), facilitates the simulations, since there is the need neither to actually compute data nor to include noise. Another important feature of the analysis of a solution in terms of resolution kernels is that it includes all the possible active sources, avoiding the enormous but never exhaustive analysis of resolution in terms of all possible pairs or triples of sources. An important claim of this work is that for the case of underdetermined systems the resolution cannot be analyzed in terms of single (or pair or triple) sources.

One-dimensional simulations

A set of 101 equally spaced electrical monopolar sources (i.e., point charges) was assumed to lie on a line, and 11 measurement electrodes were placed on a half circle above this line, as illustrated in Figure 1.

The resolution kernels for the MN, B&G, and WROP methods associated with the estimator for the activity at location A are shown in Figure 2. They are normalized such that their maximum value is equal to 1, which facilitates their comparison concerning deltaness and sidelobes. In this case, a favorable resolution kernel should have a sharp peak around location A, and fall off rapidly with increasing distance.

As seen in Figure 2, all the resolution kernels have their main peak near point A, though B&G and WROP show a slight shift. All resolution kernels tend to decrease with increasing distance to A, but clear differences can be observed: although MN has the sharpest peak around target point A, it has large negative and positive sidelobes over the whole source space. In contrast, B&G has the widest central peak and positive values everywhere. In addition, the resolution kernel of B&G decreases smoothly with increasing distance. The WROP method shows an intermediate behavior between MN and B&G. The amplitude of its sidelobes is considerably smaller in most parts of

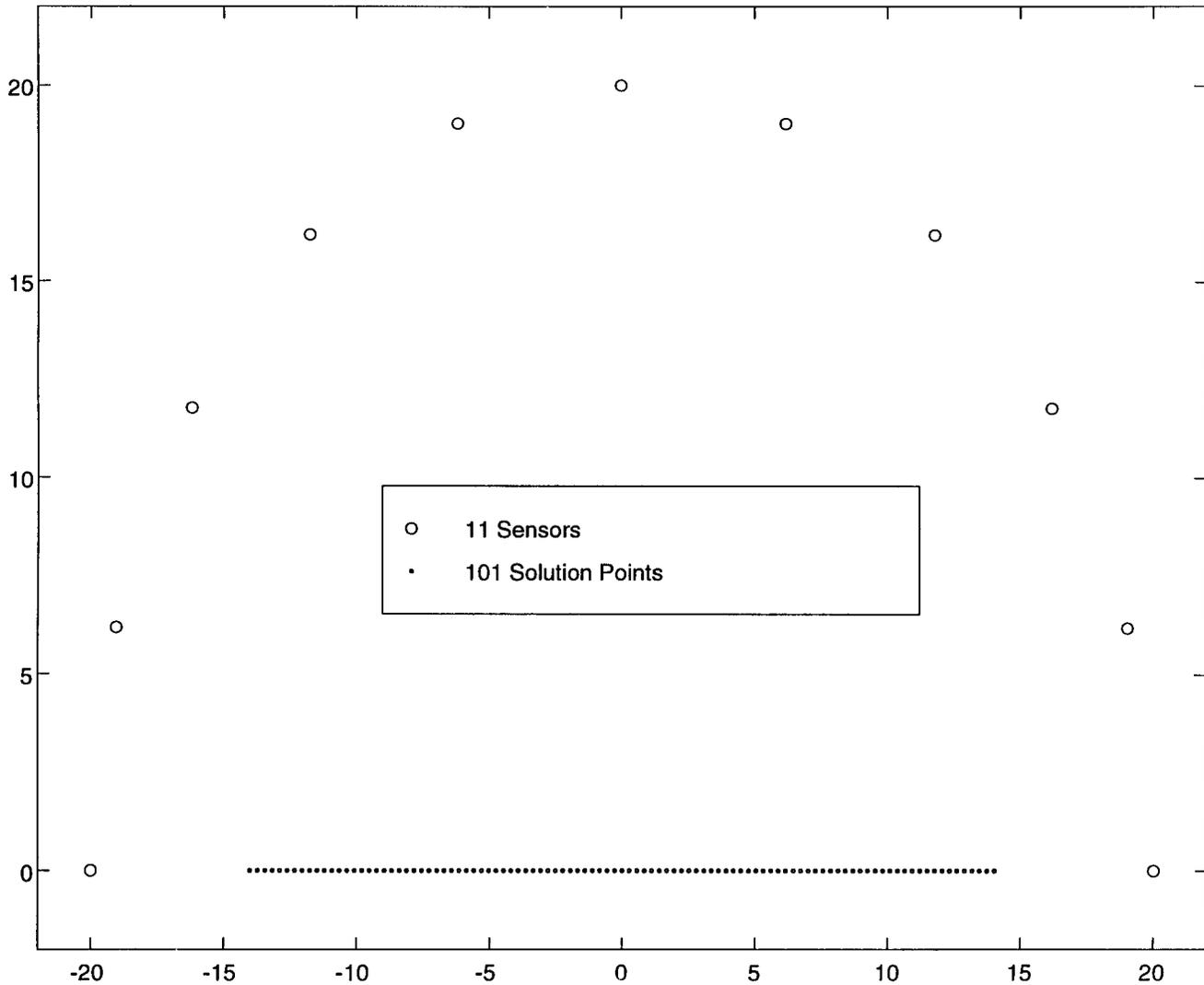


Figure 1.
Sensor and grid point configuration used in the one-dimensional simulations.

the source space than those of MN and B&G, but in contrast to B&G, these sidelobes also have negative values. Whereas the negative sidelobe on the left side of the main peak can be expected to considerably reduce the reliability of the estimate, those on the right side seem to be rather small. In fact, the resolution kernel of the WROP method is the only one which is nearly zero at larger distances from the target point.

To see the interpretation of the resolution kernels, assume that we intend to obtain an estimate for location A while minimizing the influence of other simultaneously active sources. Obviously, if a resolution kernel is not zero at A, the estimates contain information about the activity at A. If a source of

arbitrary strength exists at a location where the resolution kernel is zero, this source will have no influence on the estimator of the source at A. If a source at location B is active, the estimate for point A is a combination of the activities at locations A and B. Since all of the presented resolution kernels have greater values at A than at B, the contribution of A dominates. However, since this weight is only slightly larger, a slightly larger source at B would have the same effect on the estimate as a smaller source at A. It can thus be concluded that the activity at location A cannot be estimated independently from the activity at location B for any of the presented methods. This suggests the existence of a resolution limit at point A which no

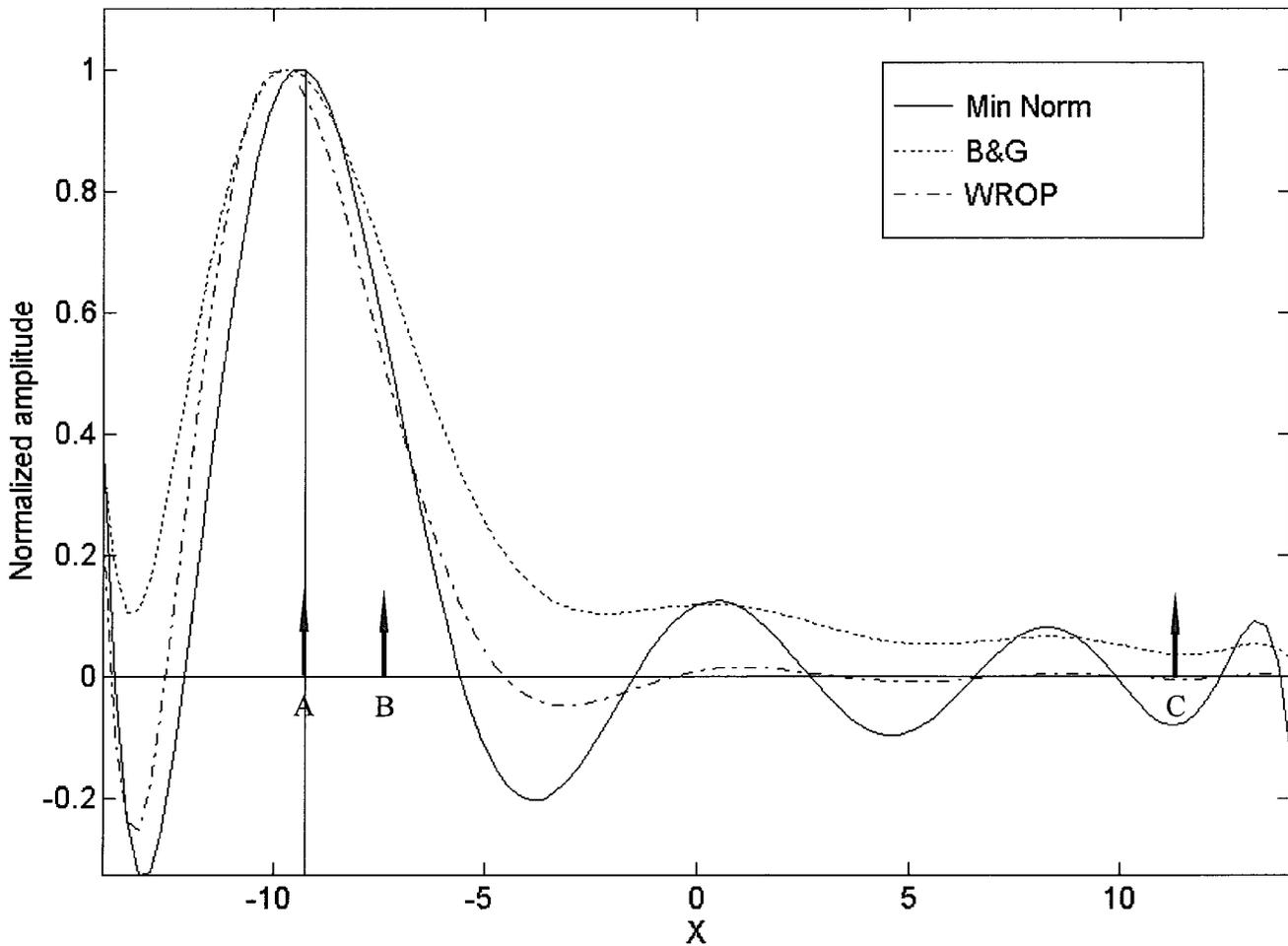


Figure 2.
Resolution kernels for the one-dimensional case associated with target point A.

linear method can overcome. This limit is closely related to the width of the peak at location A.

To analyze the differences between the estimators provided by the different methods, let us consider another source at location C. In all cases, the resolution kernels at C have considerably lower amplitude than at A. However, the MN method has a clear negative sidelobe at C, i.e., if a strong source exists at this location, it might cancel out a weaker source at A. Such situations question the interpretation of an estimate as being an average of the real source activity. In contrast, the resolution kernel of the B&G method has a positive value at C, which means that the estimate is also influenced by a source at this location, but the interpretation as a weighted average still holds. In this example, the WROP method shows the best behavior: its resolution kernel is nearly zero at C. Therefore, the activity at this location hardly contributes to the estimate of the activity in A.

It is evident from this simple 1D example that different expectations about the source lead to different opinions about which one of the given methods should be preferred. It is up to the experimenter as to which estimator is selected. Assuming one of the mathematical expressions presented above as a measure of resolution, the corresponding optimization procedure leads to the corresponding optimal estimator. In any case, the resolution kernels provide a tool for judging the properties of the resultant linear estimate [Grave de Peralta Menendez et al., 1996].

Three-dimensional simulations

Figure 3 depicts the MEG sensor configuration considered in the three-dimensional simulations, which corresponds to the 148 Biomagnetic Technologies, Inc. (BTI; San Diego, CA) whole-head system. In the coordinate system used for this and the remaining

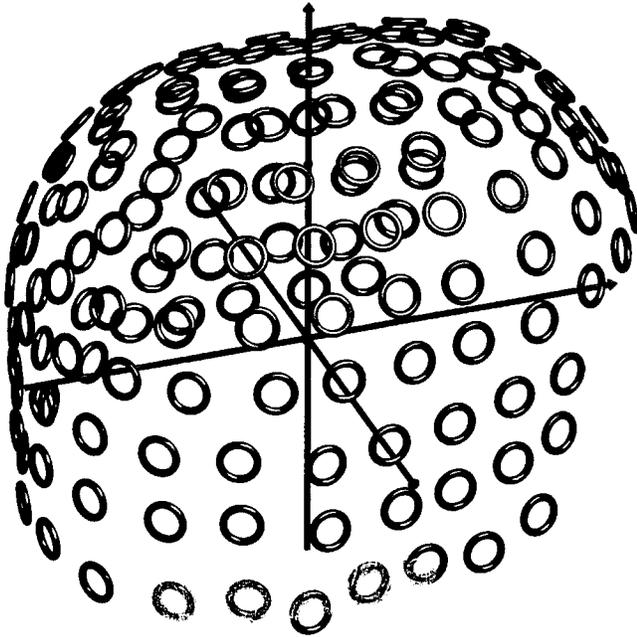


Figure 3.

Magnetic sensor configuration used for three-dimensional simulations. Magnetic field is assumed to be recorded with a 148-channel magnetometer system.

figures, the positive x-axis points to the front right part of the figure, the positive y-axis points back right, and the positive z-axis points upward (right-hand-oriented system). The selected source space is a half-sphere (in the positive z-range) of 8 cm radius, with dipoles placed on a regular grid with an intergrid distance of 1 cm. A homogeneous sphere was chosen as the volume conductor model.

Figures 4–8 show the resolution kernels obtained with the different methods for the target point, marked in the figures as a dark ball. For each figure only the resolution kernel associated with the y-component is shown. Every arrow represents how an active source at the corresponding location would affect the estimate of the y-component at the target point. A dipole perpendicular to the arrow has no influence, while a dipole parallel to the arrow has maximal influence on the estimator under consideration. The volume of the arrows is proportional to the absolute value of the resolution kernel at this location. For graphical reasons, the largest arrow was normalized to have unitary length.

Figures 4–7 show the resolution kernels obtained for a superficial source at (5, 0, 6) cm. In Figure 4 the resolution kernel of the B&G method (with the original weights proposed for the case of vector fields) is shown. As predicted in the theoretical section, the

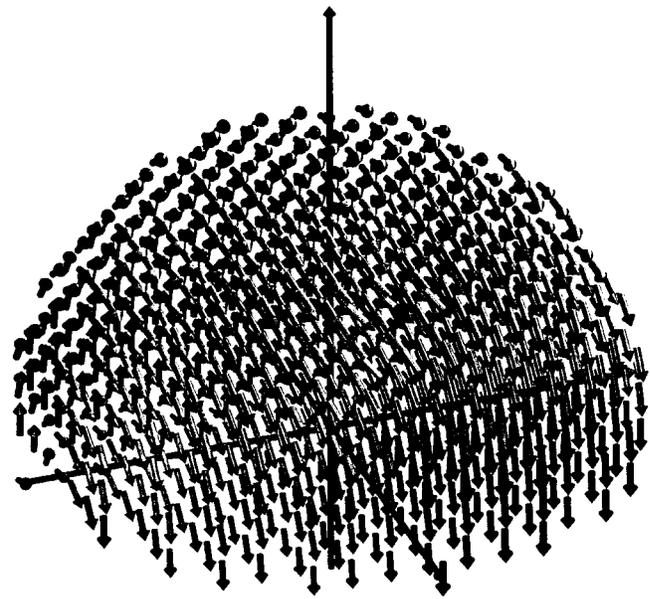


Figure 4.

Resolution kernel obtained with the B&G method and associated with the x-component of a source placed at the point represented by the dark ball. Location of the point: (0, 5, 6) cm.

resolution kernels are not decreasing with the distance to the target point, indicating a large influence of almost every dipole in the grid on the estimates for the target point. The considerable reduction of this influ-

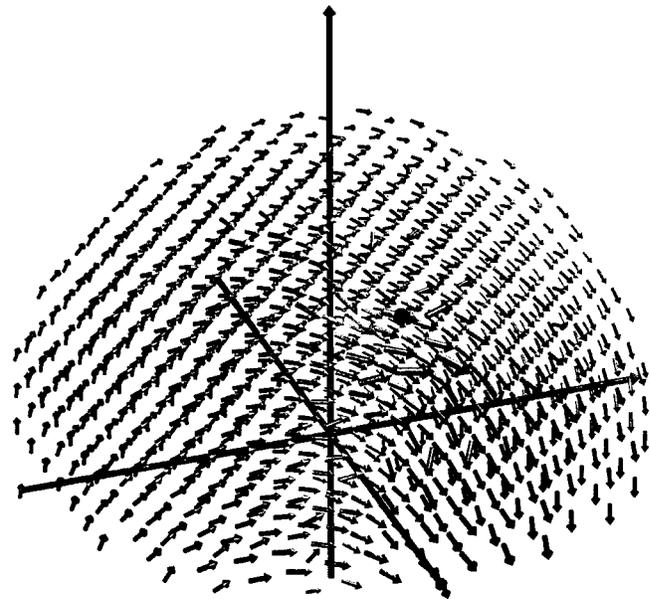


Figure 5.

Resolution kernel for the CB&G method. Target point as in Figure 4.

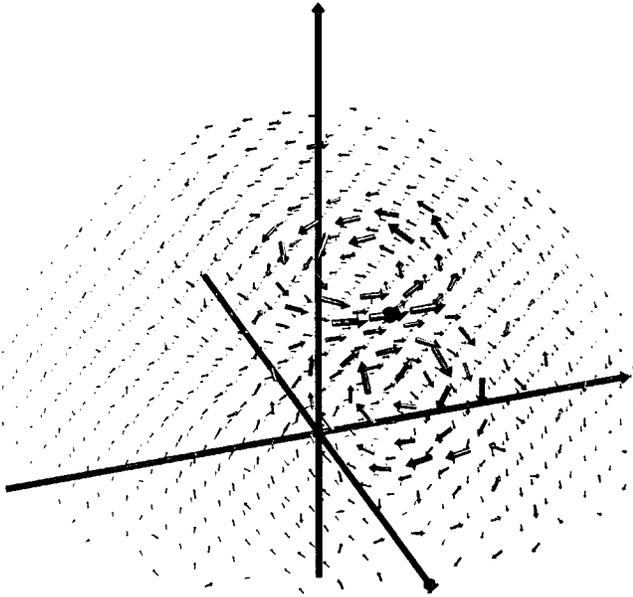


Figure 6.

Resolution kernel for the WROP method. Target point as in Figure 4.

ence, obtained by selecting the weights according to the same philosophy of Backus and Gilbert [1968] for the scalar case (CB&G), is illustrated in Figure 5. Note that the distant points have smaller arrows, which results in better estimates for the target point. The WROP method, represented in Figure 6, produces the best

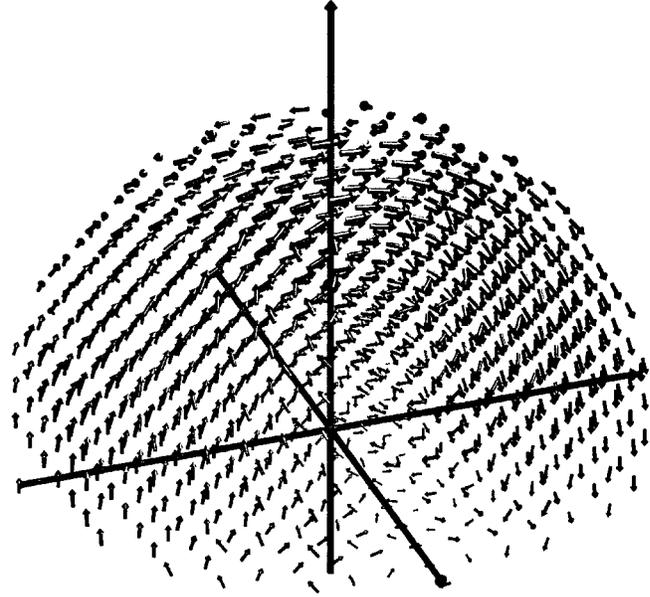


Figure 8.

Resolution kernel for the WROP method for the x-component of a source placed at a deeper location than in previous figures. Target point $(0, 0, 4)$ cm is marked by the dark ball.

estimates since the larger arrow is not only found at the target point but is also pointing in the desired y-direction. Considerable sidelobes are present, but a rough estimation of the resolving length around this point could still be derived. The resolution kernel for the MN method is shown in Figure 7. A comparison between MN and WROP shows that both resolution kernels are very similar near the target point, but that the resolution kernel of MN is more widespread. This results from the weighting strategy depending on the Euclidean distance used in the WROP method.

Figure 8 shows a resolution kernel for the WROP method, for a target point at a deeper location $(0, 0, 4)$ cm. This example illustrates the clear limits of linear inversion methods: even though the resolution was explicitly optimized for this location, the resulting estimator cannot satisfactorily suppress the influence of possible active sources, which have the same direction but are more superficially located. In fact, the resolution kernel does not even have a maximum at the location of interest and is very widespread. The interpretation of an estimate obtained with this resolution kernel as a local average of the activity around the target point is not possible. A reliable estimate for this point is only possible if no active source in nearly any superficial areas of the source space exists, a situation not likely to occur in normal brain processes. If additional superficial sources cannot be excluded a

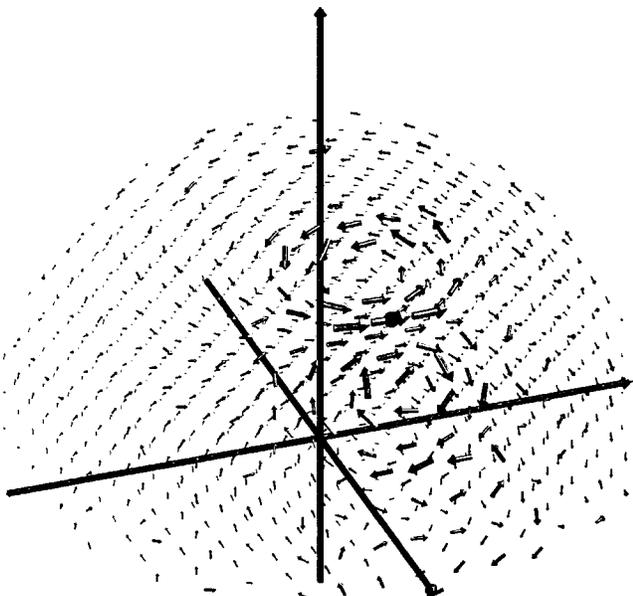


Figure 7.

Resolution kernel for the MN method. Target point as in Figure 4.

TABLE I. Strategy for selection of weights associated with some of the methods presented in the text*

Methods	Target point			Other points		
	X	Y (target)	Z	X	Y	Z
Minimum norm	α	α	α	α	α	α
Original B&G	1	0	1	1	d	1
Corrected B&G	α	0	α	$\alpha + d$	d	$\alpha + d$
WROP	$\beta + \alpha$	β	$\beta + \alpha$	$\beta + \alpha + d$	$d + \beta$	$\beta + \alpha + d$

* Target component is the y component at the target point.

priori, a reliable estimate of the activity at this deeper point is not possible.

DISCUSSION

We have shown in the theoretical section of the paper that the selection of a measure of closeness between the resolution matrix and the identity matrix results in different inverse solutions with optimal resolution kernels. The WROP method presented here is not only an improvement over the original Backus and Gilbert method for the estimation of vector fields, but is also a flexible tool for producing compact and smooth resolution kernels, as illustrated in the simulations. From a numerical point of view, this method solves a singularity that appears in the B&G method when the ideal resolution kernel is achievable and thus Equation (10) is not valid anymore. For this case of ideal resolution kernels, the singularity results in instabilities of B&G due to near-singular spread matrices. The interpretation of the WROP method in terms of a regularized minimum norm sheds light on how this problem is solved (see Equation 15). Table I illustrates the selection of weights for different methods in one example. It corresponds to the estimation of component y at the target point. The three columns under the heading Target Point represent the weights used for the target point components. The last three columns, associated with Other Points, describe the weighting strategy for all the other points, where the resolution kernel should be zero in the ideal case. The basic idea that the estimator of one component at the target point (e.g., the y component) should only depend on the y component of the other points, with decreasing influence with distance and without any influence of the other components, i.e., x and z, is clearly fulfilled in the WROP method.

One distinctive property of these solutions is their independence of the discretization of the source space. In contrast to the generalized minimum norm type methods, the estimation can be made independently for each solution point and therefore no three-dimensional grid of solution points is required. Thus, neither interpolation procedures nor artificial discretization of the solution space is needed. For the case of the generalized Wiener estimator this fact was already pointed out in Sekihara [1993]. Even when the computational burden of these methods is higher than that of the minimum norm method, it is still comparable with the burden inherent to generalized minimum norm methods that consider nondiagonal metrics [Grave de Peralta Menendez and Gonzalez Andino, 1997]. It is important to note that all these solutions are highly sensitive to variations in the selection of weights. Specifically, if the weights depend upon the geometrical distance, changes on the scale of the source space, i.e., changes in the physical units, strongly affect the shape of the resolution kernels. Also, all these solutions can be directly applied to the reconstruction of sources over the cortical surface, replacing the Euclidean distances by the appropriate distances over the cortex.

In the second part of this paper, one-dimensional simulations were presented to show the interpretation of the resolution kernels and to illustrate their use in evaluating inverse solutions with respect to the retrieval of simultaneously active sources. The same results can be used to analyze the degree of smoothing (area under the main peak) of the solutions and the existence of spurious sources (sidelobes). The resolution kernels of Figure 2 describe the typical behavior found in the simulations for all the solution points. The resolution kernels of the B&G method showed a wide peak around the target point with small or null

sidelobes; MN had a narrower peak but highly oscillating sidelobes; and WROP showed an intermediate behavior. From these resolution kernels we can infer that the estimates produced by the B&G method will generally be more blurred or smooth than those produced by MN. MN would thus show more spatial details, but the question is whether or not these details were actually present in the original source distribution or are spurious details produced by the inversion method. Considering the example in Figure 2, if any solution leads to a reconstruction with high-frequency details between points A and B, these details are probably spurious and not actually dictated by the data. This conclusion is independent of the particular data set, since the resolution kernels are intrinsically determined by the lead fields, i.e., the number and position of the sensors, and do not depend on the particular values of the measurements. To conclude: with a finite number of measurements we are unable to retrieve spatial details of the actual source distribution below a certain size. If any solution finds such details they are superfluous. If a method fails to adequately suppress or minimize such spurious features, actual and ghost sources will be mixed up in the reconstruction. Returning to the example in Figure 2, we also conclude that if any solution shows a featureless appearance between points A and B, it cannot be determined whether this is due to the actual properties of the original sources or to the smoothing imposed by the limited amount of data. However, if some details appear between points A and C, this property was likely present in the original sources, since the resolution kernel shows that features at this length scale can be adequately resolved. It follows that the spatial resolution of a solution is completely determined by the resolution kernels. It thus makes sense to design solutions with optimal resolution kernels. Nevertheless, the resolution kernels, even if optimized, can be very different from the ideal resolution kernel, i.e., peaking far from the target point and/or having highly oscillating sidelobes. For these locations it becomes very difficult to define the spatial resolution.

Concerning the 3D simulations, a discouraging but not surprising finding is that the quality of the resolution kernels degrades with the depth of the target point. A certain eccentricity value seems to exist below which all solutions fail to obtain adequately centered resolution kernels around the target point. At those points not even blurred estimates of the activity can be derived from the available data. It suggests that, unfortunately, not even a low-resolution 3D reconstruction of the current generators in the whole brain [Pascual Marqui et al., 1995] can be obtained unless

additional information about brain activity can be provided. Additional simulations for the case of electric measurements and using generalized minimum norm solutions, e.g., weighted minimum norm, maximum smoothness, and averaged solutions, showed similar results [Grave de Peralta Menendez and Gonzalez Andino, 1997]. These later simulations illustrated that while the columns of the resolution matrix might have a peak around the target point, this is not the case for their resolution kernels. In fact, both rows and columns of \mathbf{R} have nearly zero amplitude at several brain sites. Values of one in the spatial neighborhood of the main diagonal (adequate source strength estimation) were hardly ever found. In the case of near-zero amplitudes, the strength of an active source will be severely underestimated, limiting the reconstruction of simultaneously active sources of similar amplitudes but different eccentricities. These unreliable values of the source strength are one of the basic limitations of linear inverse solutions, impeding the interpretation of the reconstruction of currents as a tomography. This lack of accuracy added to the existence of ghost or spurious sources and led also to uncertainties in any posterior analysis, e.g., nonlinear analysis or correlation analysis based on these linear inverses.

We have shown that reconstruction procedures can be evaluated by exploring the resolution kernels all over the source space. Another question that can be evaluated with resolution kernels is the selection of solution points. As shown in the one- and three-dimensional simulations, the estimates obtained by any solution are, at best, averages of the real activity around the target point. Thus the intergrid distance can be selected according to the degree of smoothing of the solution. Since the estimates obtained for very close points will not produce any change in the solution, computation time can be avoided. In addition, the selection of measurement points can be assessed by selecting the optimal number and positions of the sensors to obtain "good" resolution kernels for selected brain regions.

Different aspects of these methods still require more attention, namely, the inclusion of a priori information about the sources in the design of resolution kernels and special situations where the ideal resolution kernel does not necessarily coincide with the delta function. Another point currently under study involves different alternatives to select the parameter β in the WROP method to derive several linear solutions (e.g., selecting β as λ times a matrix).

By interpreting the resolution kernels as averages of the real sources and by noting that the only averages that can be retrieved are those that can be expressed as

a linear combination of the lead field functions [Grave de Peralta Menendez and Gonzalez Andino, 1997], i.e., the rows of the lead field matrix in the discrete case, the theoretical basis for a complete analysis of resolution is available.

ACKNOWLEDGMENTS

This work was supported by Swiss National Foundation grants 4038-044081/1 and 32-37819.93 and by a grant from the Deutsche Forschungsgemeinschaft (to O.H. and H.V.).

REFERENCES

- Backus GE, Gilbert JF (1967): Numerical applications of a formalism for geophysical inverse problems. *Geophys J R Astron Soc* 13:247–276.
- Backus GE, Gilbert JF (1968): The resolving power of gross earth data. *Geophys J R Astron Soc* 16:169–205.
- Backus GE, Gilbert JF (1970): Uniqueness in the inversion of gross earth data. *Philos Trans R Soc Lond* 266:123–192.
- Bertero M, De Mol C, Pike ER (1988): Linear inverse problems with discrete data. II. Stability and regularization. *Inverse Probl* 4:573–594.
- Capon J (1969): High resolution frequency wave number spectrum analysis. *Proc IEEE* 57:1408–1419.
- Clarke CJS, Ioannides AA, Bolton JPR (1989): Localized and distributed source solutions for the biomagnetic inverse problem I. In: Williamson SJ, Hoke M, Stroink G, Kotani M (eds): *Advances in Biomagnetism*. New York: Plenum Press, pp 587–590.
- Dale AM, Sereno MI (1993): Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J Cogn Neurosci* 5:162–176.
- George JS, Aine CJ, Mosher JC, Schmidt DM, Ranken DM, Schlitt HA, Wood CC, Lewine JD, Sanders JA, Belliveau JW (1995): Mapping function in the human brain with magnetoencephalography, anatomic magnetic resonance imaging, and functional magnetic resonance imaging. *J Clin Neurophysiol* 12:406–431.
- Grave de Peralta Menendez R, Gonzalez Andino SL (1997): A critical analysis of linear inverse solutions. *IEEE Trans Biomed Eng* (in press).
- Grave de Peralta Menendez R, Gonzalez Andino S, Lütkenhöner B (1996): Figures of merit to compare linear distributed inverse solutions. *Brain Topogr* 9:117–124.
- Greenblatt RE (1993): Probabilistic reconstruction of multiple sources in the neuroelectromagnetic inverse problem. *Inverse Probl* 9:271–284.
- Hämäläinen MS, Ilmoniemi RJ (1984): Interpreting measured magnetic fields of the brain: Estimates of current distributions. Technical Report TKK-F-A559, Helsinki University of Technology.
- Hämäläinen MS, Hari R, Ilmoniemi RJ, Knuutila J, Lounasma OV (1993): Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev Mod Phys* 65:413–497.
- Hauk O (1996): Diplomarbeit am Fachbereich Physik der WWU Münster.
- Hauk O, Grave de Peralta Menendez R, Lütkenhöner B (1996): The Backus and Gilbert method and the minimum norm method applied to a simple model of a cortex fold. *Proceedings of the Third International Hans Berger Congress, Jena, Germany*.
- Ioannides AA, Bolton JPR, Hasson R, Clarke CJS (1989): Localized and distributed source solutions for the biomagnetic inverse problem II. In: Williamson SJ, Hoke M, Stroink G, Kotani M (eds): *Advances in Biomagnetism*. New York: Plenum Press, pp 587–590.
- Lütkenhöner B, Grave de Peralta Menendez R (1997): The resolution field concept. *Electroencephalogr Clin Neurophysiol* 102:326–334.
- Menke W (1989): *Geophysical Data Analysis: Discrete inverse theory*. San Diego: Academic Press.
- Pascual Marqui RD, Michel CM, Lehmann D (1995): Low resolution electromagnetic tomography: A new method for localizing electrical activity in the brain. *Int J Psychophysiol* 18:49–65.
- Penrose R (1955): A generalized inverse for matrices. *Proc Cambridge Philos Soc* 51:406–413.
- Robinson S, Rose D (1992): Current source image estimation by spatially filtered MEG. In: Hoke M, Erne S, Okada Y, Romani G (eds): *Biomagnetism: Clinical Aspects*. Amsterdam: Excerpta Medica, pp 761–765.
- Sarvas J (1987): Basic mathematical and electromagnetic concepts of the bioelectromagnetic inverse problem. *Phys Med Biol* 32:11–22.
- Sekihara K, Scholz B (1996): Generalized Wiener estimation of three dimensional current distribution from biomagnetic measurements. *IEEE Trans Biomed Eng* 43:281–291.
- Treitel S, Lines LR (1982): Linear inverse theory and deconvolution. *Geophysics* 47:1153–1159.
- Wang JZ, Williamson SJ, Kaufman L (1992): Magnetic source images determined by a lead field analysis: The unique minimum norm least square estimation. *IEEE Trans Biomed Eng* 39:665–675.