FORSCHUNGSZENTRUM JÜLICH GmbH
Zentralinstitut für Angewandte Mathematik
D-52425 Jülich, Tel. (02461) 61-6402

Interner Bericht

On Minimal l_p-Norm Solutions of the Biomagnetic Inverse Problem

*R. Beucker, H. A. Schlitt**

KFA-ZAM-IB-9614

Juni 1996
(Stand 18.06.96)

(*) Institut für Medizin (IME)

Dieser Bericht wurde zur Publikation eingereicht.

On Minimal l_p-Norm Solutions of the Biomagnetic Inverse Problem

R. Beucker* and H. A. Schlitt†

*Central Institute for Applied Mathematics
†Institute of Medicine
Research Centre Jülich GmbH
D-52425 Jülich, Germany

Abstract

In this paper we investigate the properties of minimal l_p-norm solutions to the biomagnetic inverse problem for $1 \leq p < 2$. We show that the minimal l_1-norm solutions can be interpreted as weighted l_2-norm solutions, where the weights emphasize the larger currents and make the minimal l_1-norm solutions appear more focal than the minimal l_2-norm solutions. In several examples we demonstrate that the current distribution changes continuously as the parameter p varies. Finally, a short overview about related work in robust statistics is given.

1 Introduction

In general, inverse problems occur when one wants to describe the internal property of a physical system with the help of a measured output signal. One way of interpreting the measured biomagnetic field is to solve an inverse problem for the current flow inside the brain. There are many different approaches to solving the biomagnetic inverse problem. One approach is to assume that the measured field is due to a small number of highly localized sources (current dipoles) and use a non-linear least squares fitting routine to find the location, strength and orientation of the dipoles. The major problem with dipole solutions is that a priori assumptions about the number and locations of the sources must be made.

In order to avoid making assumptions about the number and locations of the sources Hämäläinen et al. [1] proposed a minimum norm approach in which a vector field is calculated on a pre-determined grid, and each vector represents a current dipole. Generally many more source locations than measurement points are used, making

the problem highly underdetermined. Of the many current distributions which produce the same measured magnetic field, Hämäläinen et al. chose the distribution which has minimal magnitude. The solution can be determined by solving a linear least-squares problem, which is equivalent to minimizing the current distribution in the l_2-norm.

Different types of minimum norm algorithms have also been proposed where some source locations get larger weights than others. The weights can be chosen such that certain locations in the brain are preferred [2], [3], [4], [5]. The minimization can also be done in other norms such as the l_1-norm, which is defined by the sum of absolute values of the current vector components. It was observed in a limited number of examples, that current distributions calculated with the l_1-norm were much more focal than those which were calculated with the l_2-norm [6]. This is considered desirable, because one criticism of minimum norm solutions is that they are too smeared. Concentrating their interest on the l_1-norm, Matsuura et al. [7] were able to use the simplex algorithm to determine the minimum norm solution vector among all basis solutions of the underdetermined system. In this way they calculated solutions which were "maximally sparse".

Note that minimum norm methods also make assumptions about the current sources, but not the same assumptions as dipole solutions. In particular, it is assumed that sources exist only at a discrete set of locations, and these locations must be chosen a priori by the researcher. The set of locations must not be on a regular grid, and if a different set of locations is chosen, the reconstruction can change. In addition, it has not been shown that the minimal l_p-norm solution (for any p) of the current is physiologically meaningful. Nonetheless, minimum norm methods are gaining in popularity in biomagnetism and it is important to understand their properties.

In this paper we clarify why the sharpening effect appears, present an algorithm to calculate minimal l_p-norm solutions for values of p between 1 and 2, and discuss weighted l_2-norm solutions. Several examples are given to dramatically demonstrate the sharpening effect.

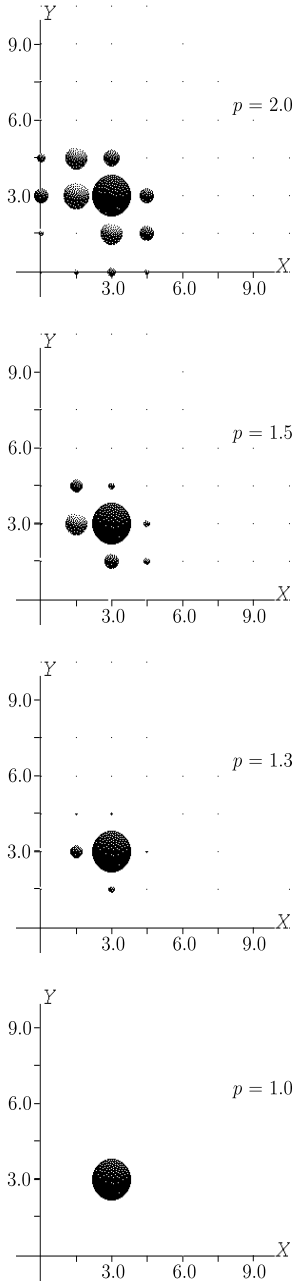

Figure 1: Minimal l_p-norm solutions for $p = 2$ (top) to $p = 1$ (bottom). The source space was a horizontal plane at $z = 3$ cm with an internode spacing of 1.5 cm. The size of the balls is proportional to the magnitude of the current at the grid points.

The paper is organized as follows. In section 2 we very briefly introduce the discrete forward model for the current distribution in a volume conductor and present a first example of the reconstructed current vector fields for different l_p-norms. In section 3, a mathematical algorithm for minimization in l_p-norms is described and an overview of related algorithms is given. We also show that minimal l_p-norm solutions for $1 < p < 2$ can be interpreted as weighted minimal l_2-norm solutions. In section 4 we show with a simple numerical example how the sharpening effect for $p \to 1$ arises. Finally, we extend the first example to show the effect of the internode spacing.

2 The Discrete Model and A First Example

For the calculation of a minimum norm solution, a discrete set of points are chosen within a pre-determined source space and the current magnitude and orientation are determined at each grid point. The assumption that sources exist only at grid points reduces the non-linear inverse problem to a linear problem. The relationship between the continuous current distribution in a volume conductor and the magnetic field is given by the Biot-Savart law. With the discrete approach using a grid one gets

$$\mathbf{b}(\mathbf{r}_j) = \frac{\mu_0}{4\pi} \sum_{i=1}^{s} \frac{\mathbf{q}(\mathbf{r}_i) \times (\mathbf{r}_j - \mathbf{r}_i)}{|\mathbf{r}_j - \mathbf{r}_i|^3}, \quad j = 1, \ldots, m. \tag{1}$$

In eq. (1), $\mathbf{q}$ denotes the moment of the current dipole, which is defined as the product of the current magnitude and orientation, s is the number of source locations, and $\mathbf{b}(\mathbf{r}_j)$ is the magnetic field at the measurement point $\mathbf{r}_j$ [8].

The m sums specify a system of linear equations, which we rewrite in matrix notation as $\mathbf{Lq} = \mathbf{b}$. With m measurement points and s grid points, $\mathbf{L}$ is a $3m \times 3s$ matrix and $\mathbf{b}$ has dimension $3m$, because the magnetic field has three components and there are current dipoles in the x, y and z directions at each of the s grid points. When only one component of the magnetic field is measured, the dimension of the system can be reduced to m rows and $3s$ columns by taking the inner product of both sides of eq. (1) with the unit direction vector at each sensor.

The most common way to select a solution of the linear system among the many possible solutions is to take the minimal l_2-norm solution. The minimal l_1-norm solution has recently generated interest because of its simple structure and the sparse reconstructions. More generally one can define a norm for a vector $\mathbf{x} \in \mathbb{R}^n$ by $\|\mathbf{x}\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, for $1 \leq p < \infty$ and $\|\mathbf{x}\|_\infty = \max_{i=1}^n |x_i|$, for $p = \infty$. For $0 < p < 1$ the expression is no longer a norm because the Minkowsky inequality, a generalization of the triangle condition, does not hold [9].

We present the first example in fig. 1 to show how extreme the differences between l_p-norm solutions of the same source can be. A single dipolar source at location

(3, 3, 3) with moment (1, −1, 0) is placed in a sphere with radius 12 cm. The magnetic field is calculated with a formula derived by Sarvas [10] for a BTi 37-channel 1st-order gradiometer array. A regular grid in the plane $z = 3$ cm with an internode spacing of 1.5 cm is placed in the sphere and the effect-matrix $\mathbf{L}$ is calculated. We calculate minimum norm solutions for different values of p between 1 and 2, using the same magnetic field and effect-matrix. The resulting vector field is plotted such that the magnitude of each direction vector is represented by the size of the ball. It is worth mentioning that for different values of p the direction of the current dipoles change. fig. 1 shows four current distributions calculated with four different values of p ($p = 2,\ p = 1.5,\ p = 1.3,\ p = 1$). In the l_2-norm solution we can see approximately 10 current dipoles with a relatively large magnitude. In the case $p = 1.3$ there are only three dipoles with a large magnitude remaining. Finally for $p = 1$ only one large dipole can be recognized. Why the reconstruction becomes sharper as p is decreased will become clearer when we have a closer look at the minimization process.

3 About l_p-Norm Minimization

The underdetermined linear system $\mathbf{Lq} = \mathbf{b}$ with rank m has no unique solution in general. One way to select a solution out of the set of non-denumerably many solutions is to solve the system by using the pseudo-inverse $\mathbf{L}^\dagger$, where the matrix $\mathbf{L}^\dagger = \mathbf{L}^T(\mathbf{L}\mathbf{L}^T)^{-1}$ [11]. This solution is equivalent to the minimization of $\|\mathbf{q}\|_2$ under the constraint $\mathbf{Lq} = \mathbf{b}$. Generalizing this approach, one can minimize $\|\mathbf{q}\|_p$ under the constraint $\mathbf{Lq} = \mathbf{b}$, where $\|\mathbf{q}\|_p$ denotes the l_p-norm of the vector $\mathbf{q}$ with $1 < p < \infty$. The l_p-norms are strictly convex, implying that there always exists a unique minimum; this is no longer true for $p = 1$ or $p = \infty$. However, as the parameter p is varied, the current distribution changes continuously [9], and under the assumption that there exist a unique minimum for the l_1-norm we can approximate the minimal l_1-norm solution by approaching p=1 from above.

In this section we show that solving the minimization problem under constraints is equivalent to calculating a minimal l_p-norm solution of an overdetermined system of linear equations. For this purpose let $n = 3s$, let $\mathbf{U}$ denote the solution space of the homogeneous linear system $\mathbf{Lq} = \mathbf{0}$, and let $k = n - m$ denote the dimension of $\mathbf{U}$. Then there exists an $n \times k$ matrix $\mathbf{A}$ with rank k such that the column space of $\mathbf{A}$ is equal to the linear subspace $\mathbf{U}$. The affine subspace of all solutions of $\mathbf{Lq} = \mathbf{b}$ can be expressed as $\mathbf{V} = \mathbf{v} + \mathbf{U}$, where $\mathbf{v}$ is any particular solution of $\mathbf{Lq} = \mathbf{b}$. We now can solve the minimization problem under constraints by minimizing $\|\mathbf{Ax} - \mathbf{v}\|_p$, where the system $\mathbf{Ax} - \mathbf{v} = \mathbf{0}$ is overdetermined.

In the case $p = 2$, the minimum is characterized by an orthogonality condition and is easily found by solving a least-squares problem. With $p \neq 2$, the orthogonality condition cannot be applied and we must use a more complicated approach. In order to simplify the notation let $\mathbf{t}(\mathbf{x}) = \mathbf{Ax} - \mathbf{v}$ be the corresponding residual vector and

let t_i denote the components of the vector $\mathbf{t}(\mathbf{x})$. Then the minimization problem $\|\mathbf{Ax} - \mathbf{v}\|_p \rightarrow \min$ is that of minimizing the function

$$f(\mathbf{x}) = \sum_{i=1}^{n} |t_i|^p. \tag{2}$$

Since the l_p-norm is strictly convex for $1 < p < \infty$, the function f has a unique minimum that can be obtained by solving the system

$$\sum_{i=1}^{n} |t_i|^{p-1} \mathrm{sgn}(t_i) a_{ij} = 0, \quad j = 1, \ldots, k, \tag{3}$$

where a_{ij} are the elements of the matrix A. The residuals, t_i, that minimize the function f in eq. (2) are also a solution to the problem of minimizing $\|\mathbf{q}\|_p$ under the constraint $\mathbf{Lq} = \mathbf{b}$, and therefore correspond to the desired current vector components.

If $p = 2$, eq. (3) are called normal equations. In case of $p > 2$ we can rewrite eq. (3) as

$$\sum_{i=1}^{n} |t_i|^{p-2} t_i a_{ij} = 0, \quad j = 1, \ldots, k. \tag{4}$$

Fletcher et al. [12] applied Newton's method to solve eq. (4). We are not very interested in this case because we want to approximate the minimal l_1-norm solution by approaching $p = 1$ from above. Assuming the residuals t_i are not zero for $i = 1, \ldots, n$ and $1 < p < 2$ eq. (3) can be rewritten as

$$\sum_{i=1}^{n} \frac{1}{|t_i|^{2-p}} t_i a_{ij} = 0, \quad j = 1, \ldots, k. \tag{5}$$

Defining $\mathbf{W} = \mathrm{diag}(\frac{1}{\sqrt{|t_1|^{2-p}}}, \ldots, \frac{1}{\sqrt{|t_n|^{2-p}}})$ eq. (5) are the normal equations of the weighted least-squares problem

$$\|\mathbf{W}(\mathbf{Ax} - \mathbf{v})\|_2 \rightarrow \min. \tag{6}$$

This formulation suggests the approach proposed by Merle and Späth [13] to solve $\|\mathbf{Ax} - \mathbf{v}\|_p \rightarrow \min$ for $1 < p < 2$ by a reweighted least-squares algorithm:

```
REWEIGHTED-LEAST-SQUARES()
 1  l ← 0
 2  for i ← 1 to n
 3      w_i^(0) ← 1
 4  repeat
 5          Determine x̂^(l) as solution of ||W^(l)(Ax − v)||_2 → min
 6          t̂ = Ax̂^(l) − v
 7          for i ← 1 to n
 8              if |t̂_i| ≥ ε
 9                 then
10                      w_i^(l+1) ← 1 / sqrt(|t̂_i|^(2−p))
11                 else
12                      w_i^(l+1) ← 1/ε
13          l ← l + 1
14  until convergence
```

Figure 2: Reweighted Least-Squares Algorithm

If a residual, $\hat{t}_i$, becomes smaller than ϵ during the iteration, the weight, $w_i^{(l+1)}$, is set to $1/\epsilon$, where ϵ is a small positive real number. We feel this is reasonable because a computer calculates with a finite numberset, which leads to roundoff errors. The iteration process is terminated if the Euclidean norm of the difference between the results, $\hat{\mathbf{x}}$, of two iteration steps is sufficiently small.

The desired current distribution corresponds to the residual vector at the final iteration step, and can be calculated from $\mathbf{t}(\hat{\mathbf{x}}) = \mathbf{A}\hat{\mathbf{x}} - \mathbf{v}$, where $\hat{\mathbf{x}}$ is the final result of the iteration process.

The local convergence of a slightly different, but more general, version of the algorithm in fig. 2 is described by Wolfe [14] in a rather involved proof. If some residuals at one iteration step are zero, he defines the weights to be zero and he then solves the remaining least-squares-problem. Merle and Späth noticed that the closer p is to 1, the more iteration steps were needed. Wolfe offers insight into the theoretical background showing under which conditions the increase in number of iteration steps happens.

There are several other approaches to solving the l_p-norm minimization problem. Ekblom [15] proposed to solve a slightly perturbed problem with a damped Newton's method. With his method it is possible to avoid zero residuals. Fischer [16] developed a globally convergent algorithm. He interprets the l_p-norm minimization as a problem with linear constraints and solves the associated dual problem with the help of lagrangian multipliers. Li [17] also presented a globally convergent algorithm, which is superlinearly convergent for $1 < p < 2$ if no component of the residual vector at the unique minimum of the function f defined in eq. (3) vanishes.

Gorodnitsky et al. [5] proposed several types of iteratively reweighted minimum norm

algorithms. Her first approach is similar to the reweighted least-squares algorithm proposed by Merle and Späth with $p = 1$, and produces similar sharpening results. Leahy et al. [18] tested Gorodnitsky's algorithm with and without noise. They found that the final error in fitting the data could be large, even for the noisless case. In their implementation they used Tikhonov regularization and the L-curve method [19] at each iteration to avoid ill-conditioning.

When noise is present for the algorithm presented in fig. 2, one can use Tikhonov regularization and the L-curve method at each iteration step as Leahy et al. did. For the examples in this paper, we calculated the basis for the solution space with the help of a truncated singular value decomposition. This guaranteed that all vectors which fulfill the constraint $\mathbf{Lq} = \mathbf{b}$ could be calculated in a numerically stable fashion.

4 Examples

The first example (fig. 1) makes it clear that current vector components with a relatively large magnitude are preferred in an l_1-norm minimization process while vector components with a smaller magnitude are damped. In other words, small residuals represent small components of the current vectors and large residuals represent large components. We now present a very simple mathematical example to demonstrate how the l_p-norms produce the sharpening effect.

4.1 Simple Mathematical Example

For this example consider the system of equations:

$$\begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 2 & 3 & 6 \end{pmatrix} \mathbf{q} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}. \tag{7}$$

We minimize $\|\mathbf{q}\|_p$ under the constraint of the linear system presented in eq. (7) using the method outlined in the previous section. We start by determining a basis of the solution space using a truncated singular value decomposition. For the calculation of a particular solution we arbitrarily chose the first and the second columns of the linear system in eq. (7). We then calculate the weights using the algorithm presented in fig. 2, and determine the residual vector $\mathbf{t}(\hat{\mathbf{x}})$ at the final iteration step. This residual vector is the desired solution to eq. (7), and is plotted in fig. 3 for a variety of p values.

We calculate the minimum norm solutions for $p = 1, 1.3, 1.5, 2$, and plot the absolute value of the vector components in fig. 3. In the case $p = 2$, the absolute value of the third component of the solution vector is only slightly larger than the other components. When p is decreased, the third component becomes bigger and the other components become smaller. In the case $p = 1$, one of the components is large

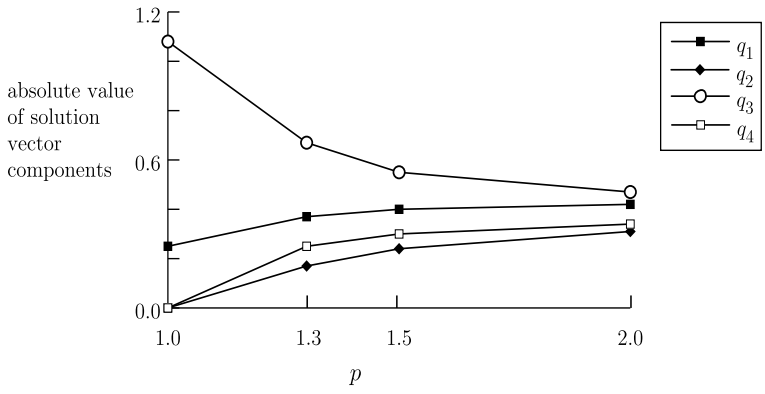

Figure 3: Solution vector components of different minimal l_p-norm solutions of the simple mathematical example (eq. (7)) for $p = 2, 1.5, 1.3, 1$

and two of the components are nearly zero, indicating a more focal, or sharper, result.

To show how large the weights are for each component, we chose $p = 1.5$ as a particular example. With 10 iteration steps we calculate the final weights (with an accuracy of 10^{-5}): $w_1^{(10)} = 1.2414$, $w_2^{(10)} = 1.3340$, $w_3^{(10)} = 1.2054$, and $w_4^{(10)} = 1.3076$. The weights for the second and the fourth components are larger than those for the first and the third components, implying that the second and the fourth components of the solution vector will be smaller. As p is decreased, this sharpening effect is strengthened.

4.2 Effect of Internode Spacing

The final example we present shows the behavior of the l_2-norm solutions in comparison to the l_1-norm solutions for different internode spacings. We return to the biomagnetic example of section 2 to investigate how different internode spacings affect the sharpness of the reconstruction.

In the example presented in fig. 4 we used the same source configuration as in section 2. In the first example we calculated l_p-norm solutions for different values of p. In this example, we use four different regular grids in the plane $z = 3$ cm with internode spacings of 1.5 cm, 2 cm, 2.5 cm and 3 cm. Note that the original source lies on a grid point only when the internode spacing is 1.5 cm and 3 cm.

The l_2-norm solution for the smallest internode spacing (1.5 cm) is smeared over the nearby grid points. As the distance between the nodes increases, the reconstruction looks more focal.

In comparison to the l_2-norm solutions the l_1-norm-solutions are much more focal for all internode spacings, although the balls with the largest magnitude always have the same position for both the l_2-case and the l_1-case. For an internode spacing of 2 cm

there are two balls; in all other cases only one large current dipole is reconstructed. In general, we find that the internode spacing has a greater effect on the l_2-norm solution than on the l_1-norm solution.

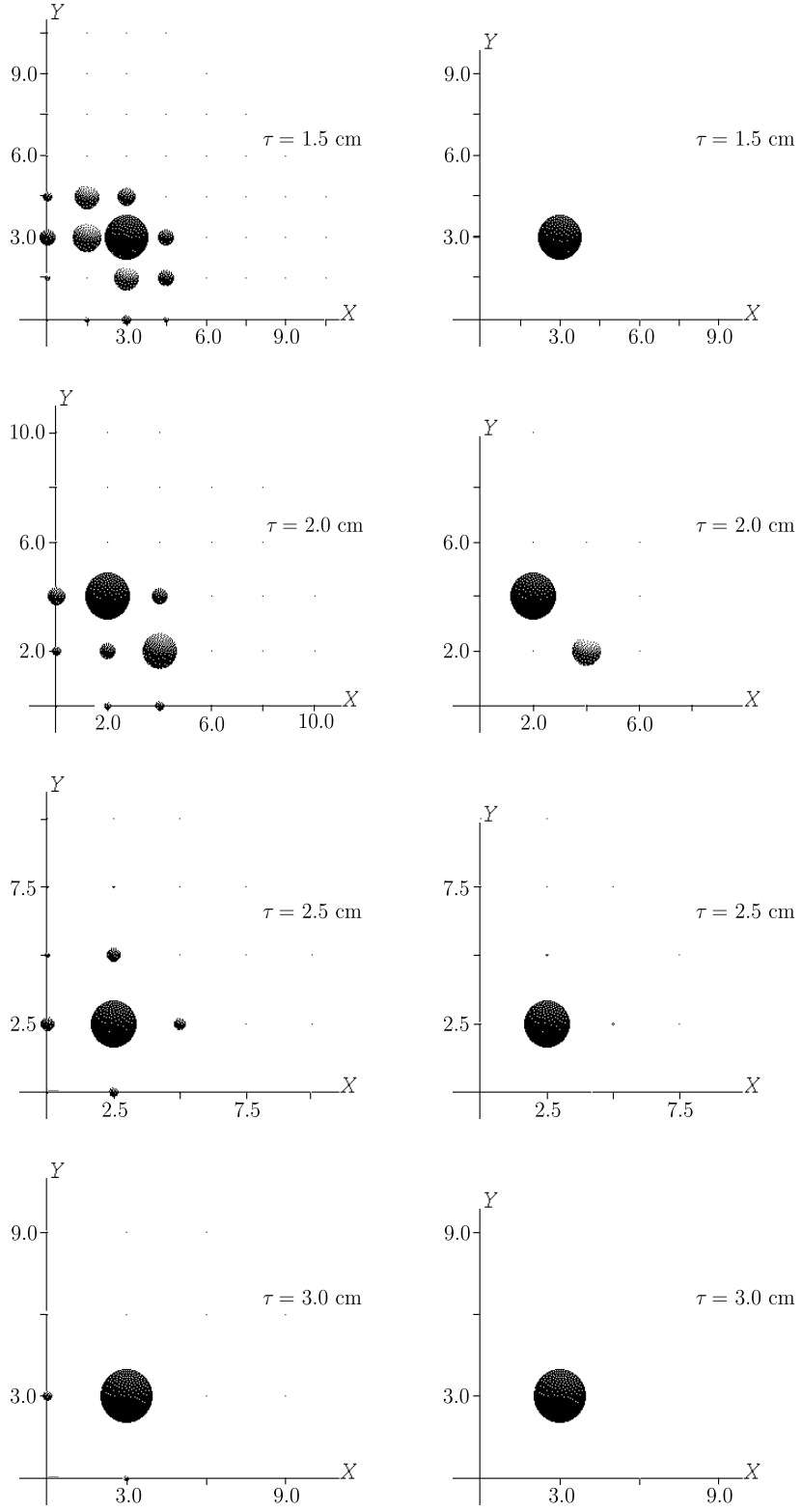

Figure 4: l_2-norm solutions (left column) and l_1-norm solutions (right column) for different internode spacings τ. The source space was a horizontal plane $z = 3$ cm. The size of the balls is proportional to the magnitude of the current at that grid point. The original source was located at (3, 3, 3) with direction (1, -1, 0).

5 Conclusions

We have seen that the l_p-norm, $1 < p < \infty$, can be interpreted as a weighted l_2-norm. Large current vector components get smaller weights than small components, so small components are damped and large ones are strengthened. This weighting makes the picture look more focal. Weighted l_1-norm minimization can partially erase this sharpening effect.

There are many similar reweighted least-squares algorithms, and many different weighting properties. The disadvantage of the algorithms is that the convergence is not clear in every case. It also can happen that there does not exist a unique minimum. The advantage of l_p-norm minimization is the uniqueness. For a detailed discussion the reader is referred to [20].

In robust statistics the l_p-norm minimization is a quite common method. Ekblom [21] compared the minimal l_p-norm solutions with Huber-estimates [22], which can be calculated with reweighted least-squares algorithms. He implemented a Monte Carlo simulation and found that Huber-estimates are superior to all pure l_p-methods with nearly normal measurement errors. For Laplace and Cauchy error distributions he proposed taking $p = 1.25$.

The remaining question is whether there exists an optimal l_p-norm for a particular class of problem. Many biomagnetism researchers have tried to answer this question by comparing different l_p-norm estimations for special examples. Linz [23] wrote (with reference to numerical regularization methods, but we find it equally applicable to the situation in biomagnetism):

> There are no general criteria by which different algorithms can be compared. Consequently, many methods are proposed which, on the evidence of a few special cases, are claimed to be effective.

6 Acknowledgments

The authors are indebted to H. M. Bücker, J. von Rango, A. Goeke, M. Sauren, P. Weidner, and E. M. E. Wermuth for giving many valuable comments and reading earlier drafts of this paper. We would like to thank H.-W. Müller-Gärtner and F. Hoßfeld for their support. We also thank H. Halling for informing us about the Matsuura paper.

References

[1] M. S. Hämäläinen and R. J. Ilmoniemi. Interpreting measured magnetic fields of the brain: estimates of current distributions. Technical Report TKK-F-A559, Helsinki University of Technology Department of Technical Physics, 1984.

[2] R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *Int. J. Psychophys.*, 18:49–65, 1994.

[3] J.-Z. Wang. Minimum-norm least-squares estimation: Magnetic source images for a spherical model head. *IEEE Trans. Biomed. Eng.*, 40:387–396, 1993.

[4] J.-Z. Wang, S. J. Williamson, and L. Kaufman. Magnetic source images determined by a lead-field analysis: The unique minimum-norm least-squares estimation. *IEEE Trans. Biomed. Eng.*, 39:665–675, 1992.

[5] I. F. Gorodnitsky, J. S. George, and B. D. Rao. Neuromagnetic source imaging with focuss: a recursive weighted minimum norm algorithm. *Electroencephalogr. Clin. Neurophysiol.*, 95:231–251, 1995.

[6] M. Wagner, H.-A. Wischmann, M. Fuchs, T. Köhler, and R. Drenckhahn. Current density reconstructions using the L1 norm. To appear in: Proceedings of the Tenth International Conference on Biomagnetism, Santa Fe, USA, February 1996.

[7] K. Matsuura and Y. Okabe. Selective minimum-norm solution of the biomagnetic inverse problem. *IEEE Trans. Biomed. Eng.*, 42:608–615, 1995.

[8] B. Jeffs, R. Leahy, and M. Singh. An evaluation of methods for neuromagnetic image reconstruction. *IEEE Trans. Biomed. Eng.*, 34:713–723, 1987.

[9] J. R. Rice. *The Approximation of Functions*. Addison-Wesley, 1964.

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

[11] J. Baumeister. *Stable Solution of Inverse Problems*. Advanced Lectures in Mathematics. Vieweg & Sohn, 1987.

[12] R. Fletcher, J. A. Grant, and M. D. Hebden. The calculation of linear best L_p approximations. *Comput J.*, 14:276–279, 1971.

[13] G. Merle and H. Späth. Computational experiences with discrete L_p-approximation. *Computing*, 12:315–321, 1974.

[14] J. M. Wolfe. On the convergence of an algorithm for discrete L_p approximation. *Numer. Math.*, 32:439–459, 1979.

[15] H. Ekblom. Calculation of linear best L_p approximation. *BIT*, 13:292–300, 1973.

[16] J. Fischer. An algorithm for discrete linear L_p approximation. *Numer. Math.*, 38:129–139, 1981.

[17] Y. Li. A globally convergent method for l_p problems. *Siam J. Optim.*, 3:609–629, 1993.

[18] R. M. Leahy, J. C. Mosher, and J. W. Phillips. A comparative study of minimum norn inverse methods for MEG imaging. To appear in: Proceedings of the Tenth International Conference on Biomagnetism, Santa Fe, USA, February 1996.

[19] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems.* Polyteknisk Forlag, 1996.

[20] H. Späth. *Mathematical Algorithms for Linear Regression.* Computer Science and Scientific Computing. Academic Press, 1991.

[21] H. Ekblom. L_p-methods for robust regression. *BIT*, 14:22–32, 1974.

[22] P. J. Huber. *Robust Statistics.* J. Wiley & Sons, 1981.

[23] P. Linz. Uncertainty in the solution of linear operator equations. *BIT*, 24:92–101, 1984.