Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 221 (2008) 386–395
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Adaptive iterative thresholding algorithms for magnetoencephalography (MEG)

Massimo Fornasier[a], Francesca Pitolli[b,*]

[a] *Program in Applied and Computational Mathematics, Princeton University, Fine Hall, Washington Road, 08544-1000 Princeton, NJ, USA*
[b] *Dipartimento di Metodi e Modelli Matematici per le Scienze Applicate, Università di Roma "La Sapienza", Via A. Scarpa 16, I-00161 Roma, Italy*

Received 16 January 2007; received in revised form 31 May 2007

Abstract

We provide fast and accurate adaptive algorithms for the spatial resolution of current densities in MEG. We assume that vector components of the current densities possess a sparse expansion with respect to preassigned wavelets. Additionally, different components may also exhibit common sparsity patterns. We model MEG as an inverse problem with joint sparsity constraints, promoting the coupling of non-vanishing components. We show how to compute solutions of the MEG linear inverse problem by iterative thresholded Landweber schemes. The resulting adaptive scheme is fast, robust, and significantly outperforms the classical Tikhonov regularization in resolving sparse current densities. Numerical examples are included.

MSC: 65J22; 65K10; 65T60; 52A41; 49M30; 68U10

Keywords: Magnetoencephalography; Inverse problems; Iterative thresholding; Adaptive algorithms; Matrix compression; Wavelets

1. Introduction

The aim of magnetoencephalography (MEG) is the analysis of brain functionality through the measurements of the tiny magnetic fields generated by neuronal currents (see, for instance, [6] and the references therein). As a matter of fact, since neuron cell functioning is mediated by electric currents, to understand brain functionality it is important to gain knowledge about the current distribution within the head. Therefore, looking from the physics side, the final goal of MEG is to accurately determine (in combination with other measurements, e.g., EEG) the current density flowing within the volume of the head in the working human brain. Among the available functional imaging techniques, such as Positron Emission Tomography (PET) or functional Magnetic Resonance Imaging (fMRI), MEG is a non-invasive technique with a high temporal resolution, typically of the order of milliseconds [6]. For this reason MEG technology appears particularly attractive. Nevertheless, the neuromagnetic field is only in the order of 10^{-13} T in magnitude, that is more than one billion times smaller than the Earth's steady magnetic field and about one million times smaller than

* Corresponding author.
E-mail addresses: mfornasi@math.princeton.edu (M. Fornasier), pitolli@dmmm.uniroma1.it (F. Pitolli).

0377-0427/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2007.10.048

the magnetic fields generated by even distant moving metal objects (e.g., cars and elevators) or electric power lines. To successfully resolve the current density flowing within the brain it is mandatory to use low noise superconducting magnetometers as well as sophisticated signal processing. In addition, the magnetic field measured externally of the head has a poor spatial resolution – at present MEG devices have at most a few hundred sensors – and can be generated by several possible current density configuration. Hence, the identification of a specific electric source configuration from the measured distribution of magnetic fields is an *ill-conditioned inverse problem*, and one of the challenging aspects of this technology. The main goal of this paper is to give an efficient and stable numerical scheme to compute the solution of the MEG inverse problem assuming that the currents flowing in the brain satisfy a sparsity constraint. In fact, according to Barlow [1], an important characteristic of sensory processing in the brain is the "redundancy reduction": the brain activity is represented as a sum of vectors of neurons, weighted by their activity, under the assumption that only a small number of neuronal areas are activated at the same time. For this reason, the sparsity assumption in the MEG model seems sufficiently realistic. Since our aim is to recover the current density, which is a vector-valued function, the joint sparsity constraint introduced in [7] to reconstruct multichannel signals seems particularly attractive.

The paper is organized as follows. In Section 2 we describe the MEG inverse problem, which involves the Biot–Savart operator, in a simplified geometry. In Section 3 we show how to transform the MEG problem into an inverse problem with joint sparsity constraints. In Section 4 we present the algorithm for its solution and prove its convergence. Section 5 is devoted to numerical implementation issues and examples.

2. The MEG inverse problem

The MEG inverse problem consists in deriving the current density $\mathbf{J}$, flowing within the brain, from the measured components of the neuromagnetic field outside the scalp.

To solve the inverse problem we need to introduce a model for the forward problem. The model is as follows.

For simplicity, we assume that the current sources are confined in a spherical region V_0, centered at the origin. Thus, the magnetic field $\mathcal{B}$ generated by the bioelectric current is given by the *Biot–Savart law*

$$\mathcal{B}(x) := \mathcal{B}(x, \mathbf{J}) = \frac{\mu_0}{4\pi} \int_{V_0} \frac{\mathbf{J}(y) \times (x - y)}{\|x - y\|^3_{\mathbb{R}^3}} \mathrm{d}(y). \tag{2.1}$$

We assume that the magnetic field is measured by sensors located on a surface $\partial\Omega$ external to the head where $\Omega \supset V_0$ is a sphere concentric to V_0 with $\delta := \mathrm{dist}(V_0, \partial\Omega) > 0$. Thus, the extracellular currents flowing in the tissue surrounding the neurons can be neglected and $\mathbf{J}$ is just the intracellular current flowing within the neurons (see [9] for details).

The *Biot–Savart operator* $\mathcal{B}$ is known to have non-trivial kernel [8] since there exist *infinitely many* current densities $\mathbf{J}$ which generate the *same* magnetic field. In addition, due to the fact that $\mathcal{B}$ is a compact operator, its generalized inverse is unbounded.

A well-known technique to stabilize the inversion is by Tikhonov regularization, i.e., by minimizing the functional

$$\mathcal{K}_\alpha(\mathbf{J}) := \|\mathcal{B}(\cdot, \mathbf{J}) - \mathcal{B}_0\|_2^2 + \alpha\|\mathbf{J}\|_2^2, \quad \alpha > 0,$$

where $\mathcal{B}_0$ is the observed magnetic field. The minimization has the unique solution

$$\mathbf{J}_\alpha := (\alpha\mathbf{I} + \mathcal{B}^*\mathcal{B})^{-1}\mathcal{B}^*\mathcal{B}_0,$$

with the property that

$$\lim_{\alpha\to 0} \mathbf{J}_\alpha \in \ker(\mathcal{B})^\perp,$$

and $\|\mathcal{B}(\cdot, \mathbf{J}_\alpha) - \mathcal{B}_0\|_2^2$ is a non-decreasing function of α. However, $\mathbf{J} \in \ker(\mathcal{B})^\perp$ cannot be a meaningful property for our problem. Therefore, the use of a simple Tikhonov regularization will risk to reconstruct incorrect current densities, while the *sparsity* assumption is more realistic.

Then we may assume that $\mathbf{J} = (J_1, J_2, J_3)^T$ is a function with compact support which is the union of few local disjoint compact volumes. By expanding the components of $\mathbf{J}$ with respect to a suitable basis of compactly supported

functions $\Psi = \{\psi_\lambda : \lambda \in \Lambda\}$ we may obtain the series

$$J_\ell = \sum_{\lambda\in\Lambda} j_\lambda^\ell \, \psi_\lambda, \quad \ell = 1, 2, 3, \tag{2.2}$$

where only *few* coefficients j_λ^ℓ are non-vanishing.

For later convenience we denote $\mathbf{j} := (j^1, j^2, j^3)^T \in \ell_2(\Lambda, \mathbb{R}^3)$, $j^\ell := (j_\lambda^\ell)_{\lambda\in\Lambda} \in \ell_2(\Lambda, \mathbb{R})$, $\ell = 1, 2, 3$. Here, the index ℓ denotes the label for the vector components, λ denotes the basis index, and in the following $\mathbf{j}_\lambda := (j_\lambda^1, j_\lambda^2, j_\lambda^3)^T \in \mathbb{R}^3$. In this situation we will say that $\mathbf{J}$ is *sparsely represented* by the basis elements Ψ. The sparsity with respect to a basis can be indeed measured directly by the ℓ_p norm for $0 < p \leq 1$: the smaller

$$\|(\mathbf{j}_\lambda)_{\lambda\in\Lambda}\|_p := \left(\sum_\lambda \left(\|\mathbf{j}_\lambda\|_{\mathbb{R}^3}\right)^p\right)^{1/p}$$

is, the fewer non-vanishing coefficients we have.

3. The MEG problem with joint sparsity constraints

Now, let us describe the MEG inverse problem with joint sparsity constraints in detail.

The data $g = (g_1, \ldots, g_N)^T \in \mathbb{R}^N$ provided by MEG are the normal components B_r of the magnetic field measured at some points $x_1, \ldots, x_N \in \partial\Omega$. If $\mathbf{e}_r(x)$ is the radially oriented unit vector at a point $x \in \partial\Omega$, then $B_r(x, \mathbf{J}) := \mathcal{B}(x, \mathbf{J}) \cdot \mathbf{e}_r(x)$ is given by

$$B_r(x) := B_r(x, \mathbf{J}) = \frac{\mu_0}{4\pi} \int_{V_0} \frac{\mathbf{J}(y) \times (x - y) \cdot \mathbf{e}_r(x)}{\|x - y\|_{\mathbb{R}^3}^3} \, \mathrm{d}(y). \tag{3.1}$$

This formula relates the radial component of the magnetic field directly to the current density $\mathbf{J}$.

For a given current density $\mathbf{J} = (J_1, J_2, J_3)^T$ we define the operator $A : L_2(V_0; \mathbb{R}^3) \to \mathbb{R}^N$ as

$$A\,\mathbf{J} = (B_r(x_1, \mathbf{J}), \ldots, B_r(x_N, \mathbf{J}))^T \in \mathbb{R}^N. \tag{3.2}$$

Now, by using in (3.1) the relation $\mathbf{v} \times \mathbf{w} \cdot \mathbf{z} = -\mathbf{z} \times \mathbf{w} \cdot \mathbf{v}$, holding for any $\mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{R}^3$, it follows that the radial component of the magnetic field at the sensor location $x_k \in \partial\Omega$ is given by

$$\begin{aligned} B_r(x_k, \mathbf{J}) &= \frac{\mu_0}{4\pi} \int_{V_0} \left(\frac{\mathbf{e}_r(x_k) \times (y - x_k)}{\|y - x_k\|_{\mathbb{R}^3}^3}\right) \cdot \mathbf{J}(y) \, \mathrm{d}(y) \\ &= \sum_{\ell=1}^3 \frac{\mu_0}{4\pi} \int_{V_0} \left(\frac{\mathbf{e}_r(x_k) \times (y - x_k)}{\|y - x_k\|_{\mathbb{R}^3}^3}\right)_\ell J_\ell(y) \, \mathrm{d}(y) \\ &= \sum_{\ell=1}^3 A_{\ell,k} \, J_\ell, \quad k = 1, \ldots, N, \end{aligned} \tag{3.3}$$

where $A_{\ell,k} : L_2(V_0; \mathbb{R}) \to \mathbb{R}$ is the operator

$$A_{\ell,k} \, f := \frac{\mu_0}{4\pi} \int_{V_0} \left(\frac{\mathbf{e}_r(x_k) \times (y - x_k)}{\|y - x_k\|_{\mathbb{R}^3}^3}\right)_\ell f(y) \, \mathrm{d}(y), \tag{3.4}$$

where $f \in L_2(V_0; \mathbb{R})$.

Due to the assumption $\delta > 0$, one can easily check that the operator A is bounded from $L_2(V_0; \mathbb{R}^3)$ to $\mathbb{R}^N$.

Let us now assume that the space $L_2(V_0; \mathbb{R})$ disposes of a stable *wavelet basis* $\{\psi_\lambda : \lambda \in \Lambda\}$ [4,13]; in particular, this means that there exist constants $C_1, C_2 > 0$ such that

$$C_1 \|f\|_{L_2(V_0;\mathbb{R})}^2 \leq \sum_{\lambda\in\Lambda} |\langle f, \psi_\lambda\rangle|^2 \leq C_2 \|f\|_{L_2(V_0;\mathbb{R})}^2 \tag{3.5}$$

for all $f \in L_2(V_0; \mathbb{R})$.

We do not enter into the details of these bases; we will just recall some of their useful properties: (i) the index $\lambda = (|\lambda|, k, e)$ encodes several different properties, respectively, the scale $|\lambda|$, the spatial location $k \in \mathbb{R}^3$, and the wavelet label e; (ii) $\Delta_\lambda := \mathrm{supp}(\psi_\lambda)$, $|\Delta_\lambda| \sim 2^{-|\lambda|}$; (iii) $\int_{V_0} y^\alpha \psi_\lambda(y)\,\mathrm{d}y = 0, \alpha = 0, \ldots, d^* \in \mathbb{N}$; (iv) $\psi_\lambda \in C^\beta(V_0)$; (v) $|\psi_\lambda| \leq C\, 2^{3/2|\lambda|}$. We refer to [13] for more details on the construction of multivariate wavelets.

Every component of a given $\mathbf{J} \in L_2(V_0; \mathbb{R}^3)$ can be decomposed in a stable way as

$$J_\ell = \sum_{\lambda \in \Lambda} j_\lambda^\ell \, \psi_\lambda, \quad \ell = 1, 2, 3. \tag{3.6}$$

Let us denote by $S : \ell_2(\Lambda; \mathbb{R}) \to L_2(V_0; \mathbb{R})$ the *synthesis operator*

$$S\,u := \sum_{\lambda \in \Lambda} u_\lambda \, \psi_\lambda, \tag{3.7}$$

where $u = (u_\lambda)_{\lambda \in \Lambda} \in \ell_2(\Lambda; \mathbb{R})$. Due to the stability inequalities (3.5) the synthesis operator S is bounded.

We define the operator $T_{\ell,k} : \ell_2(\Lambda; \mathbb{R}) \to \mathbb{R}$ as

$$T_{\ell,k} := A_{\ell,k} S, \tag{3.8}$$

so that

$$T_{\ell,k} j^\ell = \sum_{\lambda \in \Lambda} j_\lambda^\ell \left(\frac{\mu_0}{4\pi} \int_{V_0} \left(\frac{\mathbf{e}_r(x_k) \times (y - x_k)}{\|y - x_k\|^3_{\mathbb{R}^3}} \right)_\ell \psi_\lambda(y)\,\mathrm{d}(y) \right), \tag{3.9}$$

and

$$(T\mathbf{j})_k := \sum_{\ell=1}^{3} T_{\ell,k} j^\ell, \quad k = 1, \ldots, N, \tag{3.10}$$

where $\mathbf{j}$ is the sequence of the wavelet coefficients associated to the components of $\mathbf{J}$.

Thus, the MEG inverse problem can be stated as follows.

Given a set of data g we would like to compute the wavelets coefficients $\mathbf{j} \in \ell_2(\Lambda; \mathbb{R}^3)$ such that

$$T\mathbf{j} \sim g. \tag{3.11}$$

As mentioned, we can assume that $\mathbf{J} \in L_2(V_0; \mathbb{R}^3)$ is a function with compact support which is the union of few local disjoint compact volumes. Since ψ_λ are also compactly supported, only few non-vanishing coefficients $\mathbf{j}_\lambda = (j_\lambda^1, j_\lambda^2, j_\lambda^3)^T$ should be relevant to reconstruct $\mathbf{J} = S\mathbf{j}$. Additionally, different components of $\mathbf{J}$ should be assumed to be mutually correlated. A way to incorporate such correlation is the assumption of *joint sparsity* [12]. By this we mean that the pattern of non-zero coefficients representing $\mathbf{J}$ is (approximately) the same for all the components, i.e.

$$J_\ell \approx \sum_{\lambda \in \Lambda_0} j_\lambda^\ell \, \psi_\lambda, \quad \ell = 1, 2, 3, \tag{3.12}$$

where $\Lambda_0 \subset \Lambda$ is the same *finite* set for all J_ℓ's.

Thus, following [7], the solution of our problem can be modeled as the minimizer of the functional

$$\mathcal{J}^{(q)}_{\theta,\rho,\omega}(\mathbf{j}, v) := \|T\mathbf{j} - g\|^2_{\mathbb{R}^N} + \Phi^{(q)}(\mathbf{j}, v), \tag{3.13}$$

where

$$\Phi^{(q)}(\mathbf{j}, v) := \sum_{\lambda \in \Lambda} v_\lambda \, \|\mathbf{j}_\lambda\|_q + \sum_{\lambda \in \Lambda} \omega_\lambda \, \|\mathbf{j}_\lambda\|_2^2 + \sum_{\lambda \in \Lambda} \theta_\lambda (\rho_\lambda - v_\lambda)^2, \tag{3.14}$$

restricted to $v_\lambda \geq 0$. Here, $(\theta_\lambda)_{\lambda \in \Lambda}$, $(\rho_\lambda)_{\lambda \in \Lambda}$, and $(\omega_\lambda)_{\lambda \in \Lambda}$ are some suitable positive parameter sequences. The task is to minimize $\mathcal{J}^{(q)}_{\theta,\rho,\omega}(\mathbf{j}, v)$ jointly with respect to both the variables $\mathbf{j}$ and v. The first belongs to the space of signals (current densities) to be reconstructed, the second belongs to the space of sparsity indicator weights. We refer to [7]

for a detailed description of this functional. Let us only emphasize that the minimization of the term $\Phi^{(q)}(\mathbf{j}, v)$ do promote joint sparsity.

Observe that at the minimizer we will always have $0 \leq v_\lambda \leq \rho_\lambda$. Therefore, we can assume the domain of $\mathcal{J}^{(q)}_{\theta,\rho,\omega}$ to be $\ell_2(\Lambda, \mathbb{R}^3) \times \ell_{\infty,\rho^{-1}}(\Lambda)_+$, where $\ell_{\infty,\rho^{-1}}(\Lambda)_+$ denotes the (convex) cone of all non-negative sequences $(v_\lambda)_{\lambda\in\Lambda} \in \ell_{\infty,\rho^{-1}}(\Lambda)$.

4. An efficient numerical minimization

First, let us recall the convexity conditions for the functional $\mathcal{J}^{(q)}_{\theta,\rho,\omega}(\mathbf{j}, v)$. For notational simplicity let us denote $\mathcal{J}^{(q)}_{\theta,\rho,\omega}(\mathbf{j}, v)$ just by $\mathcal{J}(\mathbf{j}, v)$.

Proposition 4.1 (*[7, Prop. 2.1]*). *Let $q \in \{1, 2, \infty\}$. If $\omega_\lambda\theta_\lambda \geq \frac{\phi_q}{4}$ for all $\lambda \in \Lambda$, where $\phi_1 = 3$, and $\phi_2 = \phi_\infty = 1$, then $\mathcal{J}$ is convex. In the case of a strict inequality $\omega_\lambda\theta_\lambda > \frac{\phi_q}{4}$, $\mathcal{J}$ is strictly convex.*

An algorithm for the minimization of $\mathcal{J}(\mathbf{j}, v)$ is as follows. For some initial choice $v^{(0)}$, for example $v^{(0)} = (\rho_\lambda)_{\lambda\in\Lambda}$, the algorithm for the computation of the minimizer $(\mathbf{j}^*, v^*)$ of the functional $\mathcal{J}(\mathbf{j}, v)$ defined in (3.13) is given by

$$\begin{cases} \mathbf{j}^{(n+1)} := \arg \min\limits_{\mathbf{j}\in\ell_2(\Lambda,\mathbb{R}^3)} \mathcal{J}(\mathbf{j}, v^{(n)}) \\ v^{(n+1)} := \arg \min\limits_{v\in\ell_{\infty,\rho^{-1}}(\Lambda)_+} \mathcal{J}(\mathbf{j}^{(n+1)}, v) \end{cases} \qquad n \geq 0. \tag{4.1}$$

The minimization of $\mathcal{J}(\mathbf{j}, v^{(n-1)})$ with respect to $\mathbf{j}$ can be done by means of a fast thresholded Landweber algorithm, similar to that presented in [5]. The minimizer $v^{(n)}$ of $\mathcal{J}(\mathbf{j}^{(n)}, v)$ for fixed $\mathbf{j}^{(n)}$ can be computed explicitly. Indeed, it follows from elementary calculus that

$$v^{(n)}_\lambda = \begin{cases} \rho_\lambda - \dfrac{1}{2\theta_\lambda}\|\mathbf{j}^{(n)}{}_\lambda\|_q, & \text{if } \|\mathbf{j}^{(n)}{}_\lambda\|_q < 2\theta_\lambda\rho_\lambda, \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

From [7] we have the following result about the convergence of the above algorithm.

Theorem 4.2 (*[7, Th. 3.1]*). *Let $1 \leq q \leq \infty$ and assume that $\mathcal{J}$ is strictly convex (see Proposition 4.1). Moreover, we assume that $\ell_{2,\omega^{1/2}}(\Lambda, \mathbb{R}^3)$ is embedded into $\ell_2(\Lambda, \mathbb{R}^3)$, i.e., there exists a constant γ such that $\omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$. Then, the sequence $(\mathbf{j}^{(n)}, v^{(n)})_{n\in\mathbb{N}}$ converges to the unique minimizer $(\mathbf{j}^*, v^*) \in \ell_2(\Lambda, \mathbb{R}^3) \times \ell_{\infty,\rho^{-1}}(\Lambda)_+$ of $\mathcal{J}$. The convergence of $\mathbf{j}^{(n)}$ is weak in $\ell_2(\Lambda, \mathbb{R}^3)$ and that of $v^{(n)}$ holds componentwise.*

For the most interesting cases $q \in \{1, 2, \infty\}$, if, in addition, there exists a constant σ such that $\theta_\lambda\omega_\lambda \geq \sigma > \phi_q/4$ for all $\lambda \in \Lambda$, where $\phi_1 = 3$, $\phi_2 = \phi_\infty = 1$, then the convergence of $\mathbf{j}^{(n)}$ to $\mathbf{j}^*$ is also strong in $\ell_2(\Lambda, \mathbb{R}^3)$ and $v^{(n)} - v^*$ converges to 0 strongly in $\ell_{2,\theta}(\Lambda)$.

To implement the algorithm above we need a method to approximate the solution of the minimization with respect to the first variable $\mathbf{j}$. To this end we introduce the *thresholding operator*.

Definition 4.3. Let $1 \leq q \leq \infty$ and $\mathbf{j} \in \ell_2(\Lambda; \mathbb{R}^3)$. We define the thresholding operator by

$$\left(U^{(q)}_{v,\omega}(\mathbf{j})\right)_\lambda := (1 + \omega_\lambda)^{-1} S^{(q)}_{v_\lambda}(\mathbf{j}_\lambda),$$

where

$$S^{(q)}_v(w) = \arg \min_{z\in\mathbb{R}^3} \|z - w\|_2^2 + v\|z\|_q, \quad w \in \mathbb{R}^3. \tag{4.3}$$

The operator $S^{(q)}_v$ is given by

$$S^{(q)}_v(w) = x - P^{q'}_{v/2}(w), \tag{4.4}$$

where $P^{q'}_{v/2}$ denotes the orthogonal projection onto the norm ball of radius $v/2$ with respect to the dual norm of $\|\cdot\|_q$, i.e., the $\|\cdot\|_{q'}$-norm with q' denoting the dual index, $1/q + 1/q' = 1$.

The explicit form of $S^{(q)}_v$ for $q = 1, 2, \infty$ and $v \geq 0$ can be found in [7].

With the thresholding operator at hand the numerical scheme to compute $\mathbf{j}^{(n)}$ is given by

$$\begin{cases} \mathbf{j}^{(n,0)} & \text{given} \\ \mathbf{j}^{(n,m+1)} := U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*g - T^*T\mathbf{j}^{(n,m)}\right) & m \geq 0. \end{cases} \tag{4.5}$$

In [7, Theorem 4.9] it is shown that for a fixed $v^{(n)}$ the previous iteration is convergent in $\ell_2(\Lambda, \mathbb{R}^3)$ to the minimizer $\mathbf{j}^{(n)}$. Nevertheless, observe that each iteration involves an application of T^*T and of the thresholding operator $U^{(q)}_{v,\omega}$. The latter can be applied faster on finite sequences. Since in general T^*T is represented as an infinite matrix, its evaluation might not be exactly numerically implementable. In this paper we deal in fact with the case that Λ is not finite and, in order to realize fast and convergent schemes also in this situation, we need an algorithm for the approximate evaluation of infinite matrices. This can be done following the techniques given in [2,3,10,11]. This issue will be addressed in detail in a forthcoming paper. For now, we assume that we have the following procedure **APPLY** at our disposal:

- **APPLY**$[\varepsilon, \mathcal{N}, \mathbf{j}] \to \mathbf{j}_\varepsilon$: determines for $\mathcal{N} \in \mathcal{L}(\ell_2(\Lambda, \mathbb{R}^3))$ and for a finitely supported $\mathbf{j} \in \ell_2(\Lambda, \mathbb{R}^3)$ a finitely supported $\mathbf{j}_\varepsilon$ such that

$$\|\mathcal{N}\mathbf{j} - \mathbf{j}_\varepsilon\|_{\ell_2(\Lambda,\mathbb{R}^3)} \leq \varepsilon; \tag{4.6}$$

Moreover, in the following we can assume without loss of generality that T^*g is also a finitely supported vector. We want to substitute the exact iteration (4.5) with the following inexact one

$$\mathbf{j}^{(n,m+1)} := U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*g - \mathbf{APPLY}\left[\frac{\varepsilon^{(n,m)}}{K}, T^*T, \mathbf{j}^{(n,m)}\right]\right). \tag{4.7}$$

For a suitable choice of the approximations $\varepsilon^{(n,m)} > 0$ and for the constant $K > 0$ large enough, we derive a fully implementable and convergent scheme. In the following we assume without loss of generality that $\|T\| < 1$ which is always achievable by suitable rescaling of T and g.

Proposition 4.4. *Assume that $\omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$ and $\|T\| < 1$. Let us denote $\alpha := (1+\gamma)^{-1}(\frac{1}{K} + \|I - T^*T\|) < 1$ for $K > 0$ large enough, and*

$$\mathbf{j}^{(n,\infty)} := U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,\infty)} + T^*(g - T\mathbf{j}^{(n,\infty)})\right).$$

If $\|\mathbf{j}^{(n,\infty)} - \mathbf{j}^{(n,m)}\|_2 \leq \varepsilon^{(n,m)}$ and $\varepsilon^{(n,m+1)} := \alpha\varepsilon^{(n,m)} < \varepsilon^{(n,m)}$, then

$$\mathbf{j}^{(n,m+1)} := U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*g - \mathbf{APPLY}\left[\frac{\varepsilon^{(n,m)}}{K}, T^*T, \mathbf{j}^{(n,m)}\right]\right)$$

is such that

$$\|\mathbf{j}^{(n,\infty)} - \mathbf{j}^{(n,m+1)}\|_2 \leq \varepsilon^{(n,m+1)}. \tag{4.8}$$

Proof. By non-expansiveness of $S^{(q)}_v$ (see [7, Lemma 4.4] and its proof) we obtain

$$\begin{aligned} \|\mathbf{j}^{(n,\infty)} - \mathbf{j}^{(n,m+1)}\|_2 \leq & \left\| U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,\infty)} + T^*(g - T\mathbf{j}^{(n,\infty)})\right) - U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*(g - T\mathbf{j}^{(n,m)})\right)\right\|_2 \\ & + \left\| U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*(g - T\mathbf{j}^{(n,m)})\right) \right. \\ & \left. - U^{(q)}_{v^{(n)},\omega}\left(\mathbf{j}^{(n,m)} + T^*g - \mathbf{APPLY}\left[\frac{\varepsilon^{(n,m)}}{K}, T^*T, \mathbf{j}^{(n,m)}\right]\right)\right\|_2 \end{aligned}$$

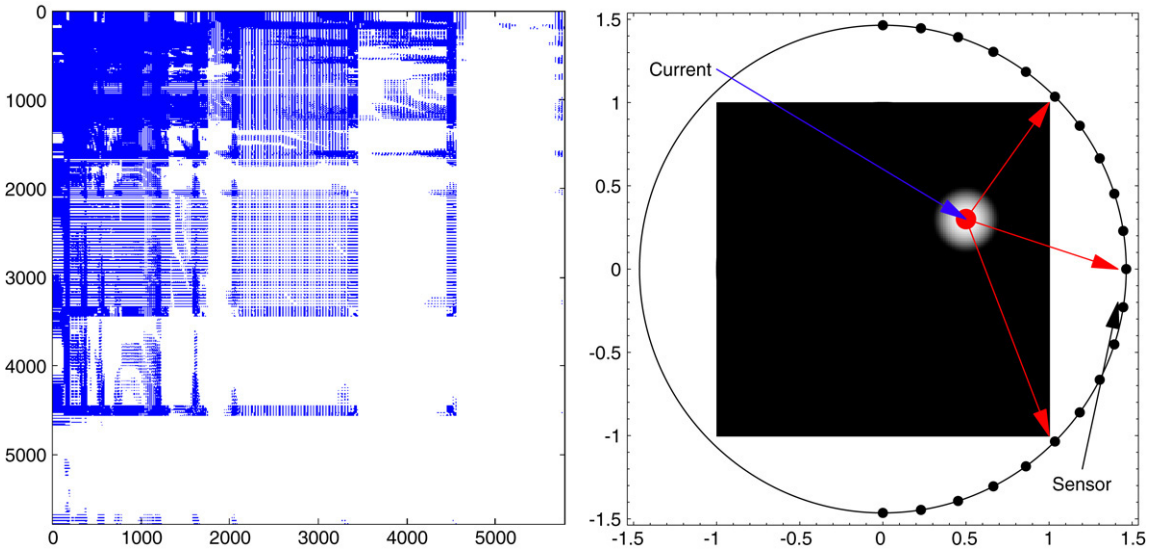

Fig. 1. **Left**: Compressed matrix T^*T associated to the Biot–Savart operator in 2D, i.e., $\mathcal{B}(x, j) := \int_{V_0} \frac{(x-y)^\perp}{\|(x-y)^\perp\|^2_{\mathbb{R}^2}} j(y)\mathrm{d}y$, with respect to Daubechies wavelets with $d^* = 4$ vanishing moments. We retain only the entries which exceed 10^{-6}. Only the 5.97% of the matrix is non-zero. **Right**: 2D model of MEG. Sensors are distributed on a semicircle around the area where current densities are distributed.

$$
\begin{aligned}
&\leq \left(\sum_{\lambda\in\Lambda} (1+\omega_\lambda)^{-2} \| S^{(q)}_{v_\lambda}(\mathbf{j}^{(n,\infty)} + T^*(g - T\mathbf{j}^{(n,\infty)})_\lambda) - S^{(q)}_{v_\lambda}(\mathbf{j}^{(n,m)} \right. \\
&\quad \left. + T^*(g - T\mathbf{j}^{(n,m)})_\lambda) \|_2^2 \right)^{1/2} + \sup_{\lambda\in\Lambda} (1+\omega_\lambda)^{-1} \frac{\varepsilon^{(n,m)}}{K} \\
&\leq \sup_{\lambda\in\Lambda} (1+\omega_\lambda)^{-1} \left(\|(I - T^*T)(\mathbf{j}^{(n,\infty)} - \mathbf{j}^{(n,m)})\|_2 + \frac{\varepsilon^{(n,m)}}{K} \right) \\
&\leq (1+\gamma)^{-1} \left(\frac{\varepsilon^{(n,m)}}{K} + \|I - T^*T\| \, \|\mathbf{j}^{(n,\infty)} - \mathbf{j}^{(n,m)}\|_2 \right) \\
&\leq \varepsilon^{(n,m+1)}.
\end{aligned}
$$

This establishes the claim. □

Let $q \in \{1, 2, \infty\}$. Assume that $\theta_\lambda \omega_\lambda \geq \sigma > \phi_q/4$ for all $\lambda \in \Lambda$, where $\phi_1 = 3$, $\phi_2 = \phi_\infty = 1$, implying that $\mathcal{J}(\mathbf{j}, v)$ is strictly convex (see Proposition 4.1). Moreover, let us assume that $\omega_\lambda \geq \gamma > 0$ for all $\lambda \in \Lambda$. Suppose $\|T\| < 1$ resulting in $\|I - T^*T\| \leq 1$. Denote

$$
\beta := \sup_{\lambda\in\Lambda} \frac{\phi_q}{4\theta_\lambda\omega_\lambda + 4\theta_\lambda(1 - \|I - T^*T\|)} \leq \frac{\phi_q}{4\sigma} < 1. \tag{4.9}
$$

With this previous result and notations, we formulate the fully implementable and convergent scheme.

Theorem 4.5. *The algorithm reads as follows. We fix the initial sequences* $\mathbf{j}^{(0)} = \mathbf{j}^{(0,0)} = 0 \in \ell_2(\Lambda, \mathbb{R}^3)$, $v^{(0)}$ *with* $0 \leq v^{(0)}_\lambda \leq \rho_\lambda$, $\varepsilon^{(0)} := \sqrt{\frac{\mathcal{J}(0, v^{(0)})}{\gamma}}$. *Under the assumptions and notations so far considered, for all* $n \in \mathbb{N}$ *we choose* $L_n \in \mathbb{N}$ *such that* $\alpha^{L_n} + \beta < 1$, *and we denote* $\mathbf{j}^{(n+1,0)} := \mathbf{j}^{(n+1)} := \mathbf{j}^{(n,L_n)}$, $\varepsilon^{(n+1,0)} := \varepsilon^{(n+1)} := (\alpha^{L_n} + \beta)\varepsilon^{(n)}$. *Then the algorithm converges strongly and in particular we have*

$$
\|\mathbf{j}^{(n,0)} - \mathbf{j}^*\|_2 \leq \varepsilon^{(n)}.
$$

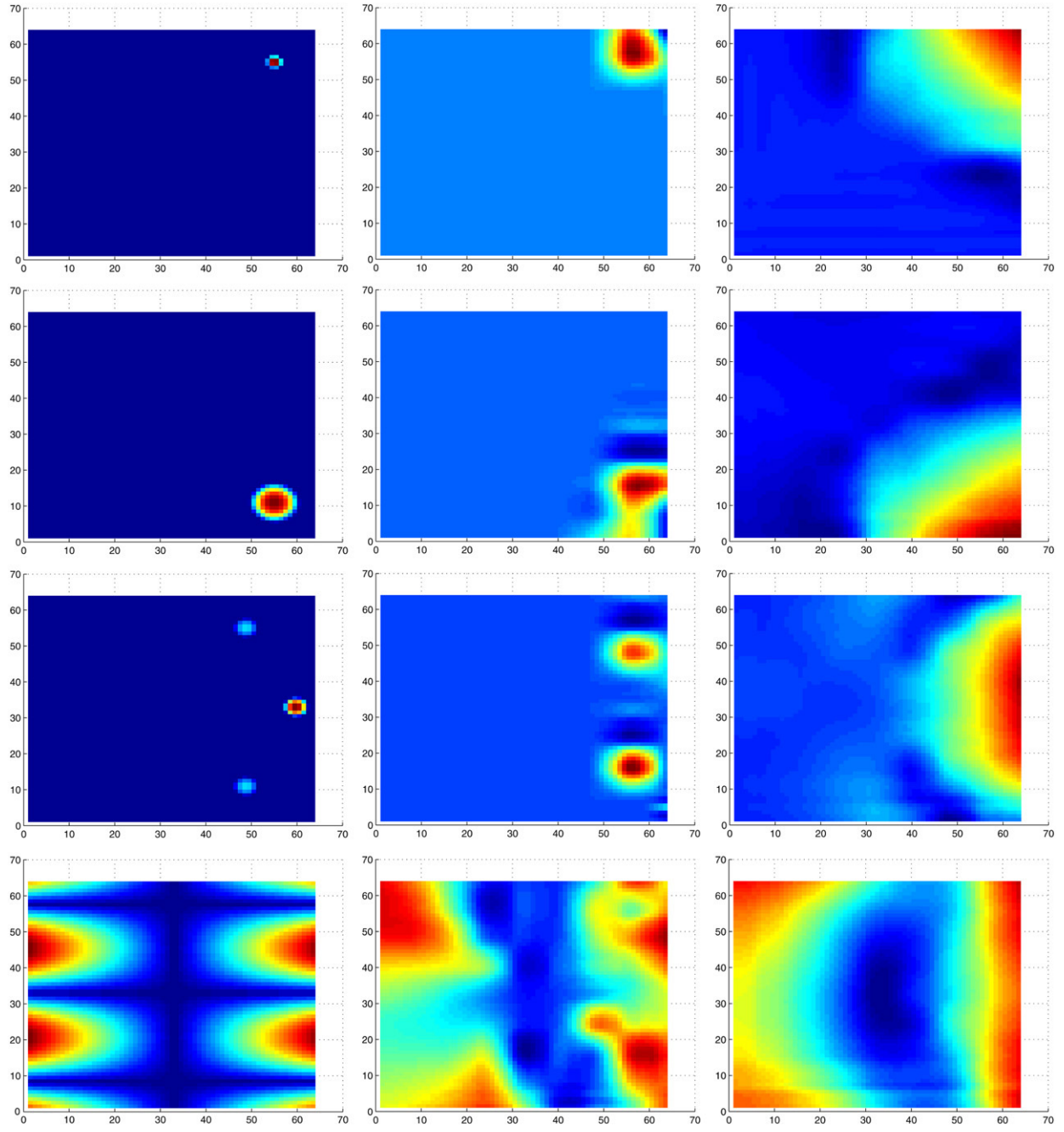

Fig. 2. In the first column we illustrate some fictitious spatially localized (sparse) current densities as in the 2D model of Fig. 1. In the second column we illustrate the result of the reconstruction due to our algorithm for suitable choices of the parameters. The third column shows the corresponding reconstruction due to Tikhonov regularization. The use of **JOINTSPARSE** significantly outperforms the classical Tikhonov regularization, improving the spatial resolution of the current densities.

Proof. We prove the statement by induction. We have

$$\|\mathbf{j}^{(0,0)} - \mathbf{j}^*\|_2^2 = \|\mathbf{j}^*\|_2^2 \leq \frac{\sum_\lambda \omega_\lambda \|\mathbf{j}_\lambda^*\|_2^2}{\gamma}$$

$$\leq \frac{\mathcal{J}(\mathbf{j}^*, v^*)}{\gamma} \leq \frac{\mathcal{J}(0, v^{(0)})}{\gamma} = (\varepsilon^{(0)})^2.$$

Assume then that $\|\mathbf{j}^{(n-1,0)} - \mathbf{j}^*\|_2 \leq \varepsilon^{(n-1)}$.

From Proposition 4.4 and [7, Proposition 5.4] we have

$$\begin{aligned}\|\mathbf{j}^{(n,0)} - \mathbf{j}^*\|_2 &\leq \|\mathbf{j}^{(n,0)} - \mathbf{j}^{(n-1,\infty)}\|_2 + \|\mathbf{j}^{(n-1,\infty)} - \mathbf{j}^*\|_2 \\ &\leq \alpha^{L_{n-1}}\varepsilon^{(n-1)} + \beta\|\mathbf{j}^{(n-1,0)} - \mathbf{j}^*\|_2 \\ &\leq \left(\alpha^{L_{n-1}} + \beta\right)\varepsilon^{(n-1)} = \varepsilon^{(n)}.\end{aligned}$$

This concludes the proof. □

5. Numerical implementation

A straightforward computation shows that

$$\left(T^*T\,\mathbf{j}\right)_{\lambda,\ell} = \sum_{\mu\in\Lambda}\sum_{m=1}^{3}\left(\sum_{k=1}^{N}(A_{\ell,k}\,\psi_\lambda)(A_{m,k}\psi_\mu)\right)j_\mu^m, \quad \lambda \in \Lambda,\ \ell = 1,2,3. \tag{5.1}$$

Let us denote by $\mathcal{M}$ the matrix associated to T^*T in wavelet coordinates whose entries are $(\mathcal{M})_{(\lambda,\ell),(\mu,m)} := (\sum_{k=1}^{N}(A_{\ell,k}\psi_\lambda)(A_{m,k}\psi_\mu))$. With similar arguments as in [10, Lemma 8.1] we can show that

$$|\mathcal{M}_{(\lambda,\ell),(\mu,m)}| \leq C2^{-(|\lambda|+|\mu|)(3/2+d^*+1)} \times \begin{cases} N\,\mathrm{dist}(\Omega_\lambda, \Omega_\mu)^{-(d^*+3)} & \mathrm{dist}(\Omega_\lambda, \Omega_\mu) > 0 \\ \displaystyle\sum_{k=1}^{N}(\mathrm{dist}(x_k, \Omega_\lambda)\,\mathrm{dist}(x_k, \Omega_\mu))^{-(d^*+3)} & \text{otherwise,}\end{cases}$$

where d^* is the number of vanishing moments of the wavelets. These latter estimates allow for the formulation of an efficient **APPLY** routine (compare [10, Section 8]). In Fig. 1 we show the compressibility properties of the matrix $\mathcal{M}$ for the corresponding 2D model of the Biot–Savart inversion. Fig. 2 shows some numerical results and the comparison with the classical Tikhonov regularization. The 2D model corresponds to the 3D situation where the current densities are furthermore constrained to be vectors oriented along the orthogonal axis to the plane where the 2D magnetic field lies on, $\mathbf{J}(x_1, x_2, x_3) = (0, 0, j(x_1, x_2))$. Of course, this model does not represent a realistic situation. However, despite its simplicity, it exhibits all the interesting mathematical and numerical features of the more general 3D case, and it is simple to implement.

Acknowledgements

Massimo Fornasier acknowledges the financial support provided by the European Union's Human Potential Programme under contract MOIF-CT-2006-039438. He also thanks Ingrid Daubechies and Ignace Loris for the fruitful discussions on sparse recovery during the preparation of this work.

References

[1] H.B. Barlow, What is the computational goal of the neocortex? in: Large-scale Neuronal Theories and the Brain, MIT Press, Cambridge, MA, 1994.
[2] S. Dahlke, M. Fornasier, T. Raasch, Adaptive frame methods for elliptic operator equations, Adv. Comput. Math. 27 (1) (2007) 27–63.
[3] S. Dahlke, T. Raasch, M. Werner, M. Fornasier, R. Stevenson, Adaptive frame methods for elliptic operator equations: The steepest descent approach, IMA J. Numer. Anal. 27 (2007) 717–740.
[4] I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.
[5] I. Daubechies, M. Defrise, C. DeMol, An iterative thresholding algorithm for linear inverse problems, Comm. Pure Appl. Math. 57 (1994) 1413–1457.
[6] C. Del Gratta, V. Pizzella, F. Tecchio, G.L. Romani, Magnetoencephalography—a noninvasive brain imaging method with 1ms time resolution, Rep. Progr. Phys. 64 (2001) 1759–1814.

[7] M. Fornasier, H. Rauhut, Recovery algorithms for vector valued data with joint sparsity constraints, SIAM J. Numer. Anal. (in press).
[8] L. Kühn, Magnetic Tomography. On the Nullspace of the Biot–Savart Operator and Point Sources for Field and Domain Reconstruction, Georg-August-Universität Göttingen, Germany, 2005.
[9] J. Sarvas, Basic mathematical and electromagnetic concepts of the biomagnetic inverse problem, Phys. Med. Biol. 32 (1987) 11–22.
[10] R. Schneider, Multiskalen- und Wavelet-Matrixkompression: Analysisbasierte Methoden zur effizienten Lösung grosser vollbesetzter Gleichungssysteme, in: Advances in Numerical Mathematics, Teubner, Stuttgart, 1998.
[11] R. Stevenson, Adaptive solution of operator equations using wavelet frames, SIAM J. Numer. Anal. 41 (2003) 1074–1100.
[12] J. Tropp, Algorithms for simultaneous sparse approximation. Part II: Convex relaxation, Signal Process. 86 (2006) 589–602.
[13] K. Urban, Wavelets in Numerical Simulation. Problem Adapted Construction and Applications, Springer-Verlag, 2002.