RECONSTRUCTION OF FUNCTIONAL ACTIVATIONS IN DIFFUSE OPTICAL IMAGING

Mathews Jacob, Vlad Toronov, Yoram Bresler

University of Illinois at Urbana Champaign

Xiaofeng Zhang, Andrew Webb

Penn State University

ABSTRACT

We propose a new algorithm for the estimation of functional activations in diffuse optical imaging. Our approach considers the activations to be support limited. We simultaneously estimate the function values as well as the support from the available measurements. Since this scheme exploit the structure inherent to functional imaging, it provide reconstructions with better spatial resolution and is more robust to noise.

I. INTRODUCTION

High temporal resolution and sensitivity to oxy and de-oxy heamoglobin concentrations in diffuse optical imaging(DOI) makes it an attractive technique for functional brain imaging [1], [2]. However, its main drawback is the limited spatial resolution, mainly due to the diffuse nature of light propagation, few source detector pairs, noisy measurements, and limited penetration of light into the brain.

The classical linear reconstruction techniques are currently the widely used schemes in DOI [3]. These techniques being general, are not efficient to exploit the structure inherent to the problem. The main focus of this paper is to develop a new reconstruction algorithm to improve the spatial resolution as well the robustness to noise. Since our approach utilize the sparsity, the connectedness of the activations, and the smooth nature of activations over the support, we obtain more robust reconstructions with improved spatial resolution. In traditional schemes, the reconstruction is made well posed by restricting the solution to a low-dimensional eigenspace of the system matrix. This space, do not represent the class of natural images, in general, and hence may result in artifacts. In comparison, the new method enforce constraints that are expected in a natural setting to obtain better-posed reconstructions.

Since we have to enforce extra constraints in addition to sparsity, the standard l_1 minimization approach to solve sparse problems [4] cannot be used in our setting. We propose an algorithm based on variational principles, which is an extension of [5]. We consider the support of the activated regions and the function values over this support as unknowns and solve for them using a two-step iterative scheme. In the first step, we assume an initial support and solve for the activations over this region using the conjugate gradients optimization algorithm. In the next step, we evolve the support so as to minimize an appropriate cost function.

This work was supported by the Beckman Foundation and NIH grant 5R01MH065429-2.

II. PRELIMINARIES

In this section, we formulate the forward problem. Assuming the perturbation of the absorption coefficient (denoted by $\triangle\mu_a(\mathbf{x})$) to be small, the corresponding perturbation in the optical fluence at the j^{th}detector due to the i^{th} source is

$$\triangle\phi(i,j) = \int_{\mathbb{R}^3} \kappa_{i,j}(\mathbf{x})\triangle\mu_a(\mathbf{x})d\mathbf{x}. \tag{1}$$

The sensitivity functions $\kappa_{i,j}(\mathbf{x})$ are termed as light bundles in the diffuse optical imaging (DOI) literature. It is experimentally observed that the measurement noise is a zero-mean Gaussian process with standard deviation proportional to the amplitude of the baseline signal. Collecting the measurements corresponding to the different source-detector pairs into a single vector, and denoting $f(\mathbf{x}) = \triangle\mu_a(\mathbf{x})$, we obtain

$$\mathbf{y} = \int_{\mathbb{R}^3} \boldsymbol{\kappa}(\mathbf{x})f(\mathbf{x})d\mathbf{x} + \boldsymbol{\eta}. \tag{2}$$

The main issue in the inversion of this linear problem is the limited number of source detector pairs, the large number of unknowns, and the noisy nature of the measurements.

III. RECONSTRUCTION

By collecting the voxel values of $f(\mathbf{x})$ into a vector $\mathbf{f}$, the forward model (3) can be expressed as a matrix equation $\mathbf{y} = \mathbf{A}\mathbf{f} + \boldsymbol{\eta}$. Each voxel in the reconstructed image constitutes a column of $\mathbf{A}$. Similarly, a measurement corresponds to a row in $\mathbf{A}$. In typical diffuse optical imaging (DOI) measurements, the number of columns are far greater than that of the rows, resulting in an ill-posed problem.

Similar ill-posed problems were considered in the context of sparse sampling theory, where it is shown that a sparse signal can be recovered uniquely from few measurements. It is shown in [4], [6] that if the sub-matrix of $\mathbf{A}$, obtained by selecting any of its M columns is invertible, an activation pattern with at-most M non-zero voxels can be uniquely reconstructed with probability one. Unfortunately, the maximum M, such that the condition number of every $M \times M$ sub-matrix of $\mathbf{A}$ is reasonably low, is fairly small

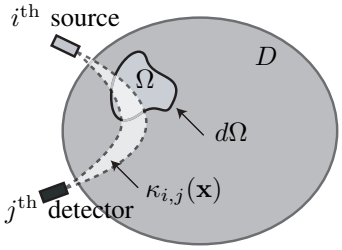


Fig. 1. The setup of the reconstruction algorithm illustrated in 2-D. Ω denotes the region of activation (where the function f is assumed to be non-zero). $d\Omega$ is the boundary of Ω and $\mathcal{N}(x)$ is the unit outward normal to the boundary. D denotes mask of the brain region, and $\kappa_{i,j}(\mathbf{x})$ is the light bundle. With this assumption of f being zero in $D \setminus \Omega$, the forward model can be formulated as (3).

for the DOI forward model ($M <= 16$). Thus sparsity alone cannot ensure the uniqueness of the reconstructions. We propose to enforce additional constraints that are relevant to the functional imaging context (eg. active regions being connected, the boundary of the activations being smooth, the activation pattern being smooth) to further reduce the degrees of freedom. With these added constraints, we would be searching for an activation pattern in the class of reconstructions that satisfy these constraints; we hypothesize that these constraints along with sparsity will ensure well-posed reconstructions.

Since we have to enforce additional constraints, it is not possible to use standard algorithms for sparse optimization like linear programming. Hence, we propose an alternate algorithm using variational principles. Assuming the activations to be support limited to subregions of the brain specified by $\Omega \subseteq D$ (see Fig. 1), we express the forward model as

$$\mathbf{y} = \int_{\mathbb{R}^3} \boldsymbol{\kappa}_\Omega(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} + \boldsymbol{\eta}, \tag{3}$$

where

$$\boldsymbol{\kappa}_\Omega(\mathbf{x}) = \begin{cases} \boldsymbol{\kappa}_\Omega(\mathbf{x}) & \text{if } \mathbf{x} \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

In the above equation, both the function values and the support of f in (3) are unknowns. We pose the estimation of these quantities as a numerical optimization problem, where the cost function is

$$\begin{aligned} \mathcal{C}(f,\Omega) &= \|\mathbf{y} - \boldsymbol{\kappa}_\Omega f\|^2_{\mathbf{W}} + \lambda \int_\Omega |\nabla f|^2 d\mathbf{x} \\ &\quad + \mu \int_\Omega |f|^2 d\mathbf{x} + \nu \int_\Omega d\mathbf{x} \end{aligned} \tag{5}$$

The first term in the above equation is a measure of the data consistency. $\|\|_{\mathbf{W}}$ denotes the weighted l_2 norm: $\|\mathbf{y}\|^2_{\mathbf{W}} = \|\mathbf{W}\mathbf{y}\|^2$. We choose the weighting matrix $\mathbf{W} = \text{diag}(1/\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ is the baseline signal vector to whiten the noise process. We now simplify this weighted l_2 norm to the standard l_2 norm by considering $\mathbf{y}' = \mathbf{W}\mathbf{y}$ and $\boldsymbol{\kappa}' = \mathbf{W}\boldsymbol{\kappa}$. For the rest of the paper, we make this assumption. However, we will just denote $\mathbf{y}'$ and $\boldsymbol{\kappa}'$ by $\mathbf{y}$ and $\boldsymbol{\kappa}$ respectively.

The second and third terms in (5) are the standard Tikhonov regularization terms. As $\lambda \to \infty$, ∇f would only be supported on $d\Omega$, the boundary of Ω. In this case, we would be reconstructing a piecewise constant function f. The last term in (5) constrains the volume of the activations. It ensures that the estimation of the support Ω is well-posed. The parameters λ, μ and ν control the strength of the additional constraints relative to the data consistency constraint. Note that in the absence of the second and third terms, the cost function is an l_0 minimization problem; one tries to find an f with the smallest support.

The reconstruction of the activation pattern thus reduces to the minimization of (5): $f^* = \arg\min_{f,\Omega} \mathcal{C}(f,\Omega)$. We use a two-step alternating minimization algorithm to solve this problem. In the first step, we estimate the optimal f, assuming Ω to be known. In the next step, we update Ω, assuming the value of f from the previous iteration. The derivation of the optimal f given Ω can be formulated as

$$f^* = \arg\min_f \left[\|\mathbf{y} - \boldsymbol{\kappa}_\Omega f\|^2 + \mu \int_\Omega |f|^2 d\mathbf{x} + \lambda \int_\Omega |\nabla f|^2 d\mathbf{x} \right] \tag{6}$$

Note that we omitted the last term in (5) since it is independent of f. Using variational principles, we show that the solution to this problem satisfy the linear conditions:

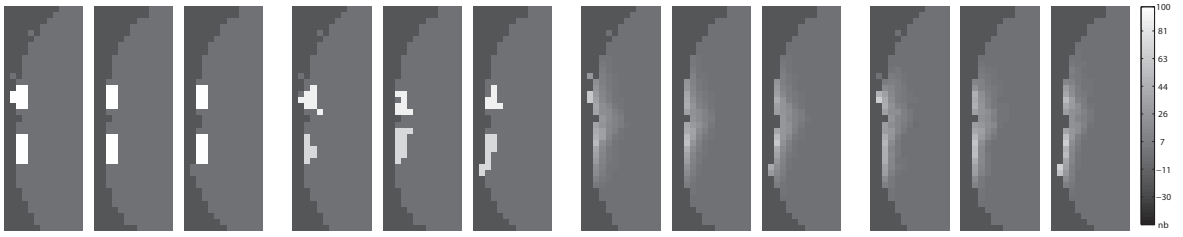
$$\begin{aligned} \left(\boldsymbol{\kappa}_\Omega^H \boldsymbol{\kappa}_\Omega\right) f + \mu f - \lambda \nabla^2 f &= \boldsymbol{\kappa}_\Omega{}^* \, \mathbf{y}; \text{ on } \Omega \\ \nabla f \cdot \mathcal{N} &= 0; \text{ on } d\Omega, \end{aligned} \tag{7}$$

where $\mathcal{N}$ denote the unit normal of the surface $d\Omega$ and $\mathbf{v}_1.\mathbf{v}_2$ denote the dot-product between $\mathbf{v}_1$ and $\mathbf{v}_2$. The operator $\boldsymbol{\kappa}^H$ denotes the adjoint of $\boldsymbol{\kappa}$. Assuming a discrete image model and finite difference approximation of the derivative operator, (7) can be reformulated as an $\mathbf{A}\mathbf{x} = \mathbf{b}$ problem. We use conjugate gradients algorithm to solve for the optimal f.

In the second step, we update the current estimate of Ω, assuming the value of f to be known. The standard procedure is to evolve the boundary $d\Omega$ with a specified speed $v(\mathbf{x})$ along the unit outward normal of $d\Omega$ (denoted by $\mathcal{N}$). The speed function is chosen such that the cost function (5) is minimized; We show that the speed function that optimally minimize the cost is given by

$$v_{\text{opt}}(\mathbf{x}) = \mathbb{R}e\left(-2f^* \, \boldsymbol{\kappa}_\Omega^H \left(\boldsymbol{\kappa}_\Omega f - \mathbf{y}\right) + \mu |f|^2 + \lambda |\nabla f|^2 + \nu \right) \tag{8}$$

Note that this scheme is equivalent to minimizing $\mathcal{C}$ with respect to Ω using a steepest descend algorithm. We use the level-set scheme to represent the region boundaries due to its efficiency in dealing with volumes of arbitrary topology. The boundary $d\Omega_t$ at the time instant t is represented as the



(a) Original (b) Levelset (c) SIRT (d) Conjugate gradients

Fig. 2. Reconstruction in the presence of noise. The range of the original signal is scaled to be in the range $0-100$; the same scale factor is applied to the reconstructions as well. Note that the SIRT and CG give smooth reconstructions. Also note that their amplitudes (from the colorbar) are not in good agreement with the original. On the other hand the new algorithm gave reasonable reconstructions indicating its improved resolving capability. The reconstructions took approximately 51s on a Pentium 4, 3.2 GHz computer.

zero level set of a 3-D function $\phi_t(\mathbf{x}) : \mathbb{R}^3 \rightarrow \mathbb{R}$: $d\Omega = \{\mathbf{x} \in \mathbb{R}^3 |\ \phi_t(\mathbf{x}) = 0\}$. The support of f is given by

$$\Omega = \{\mathbf{x} \in \mathbb{R}^3 |\ \phi_t(\mathbf{x}) > 0\} \tag{9}$$

In the level-set framework, the deformation of the region from Ω_t to Ω_{t+1}, by evolving the boundary with a velocity $v(\mathbf{x})$, is obtained by solving the partial differential equation

$$\frac{\partial \phi}{\partial t} = v\,|\nabla \phi| \tag{10}$$

If the boundary flow, specified by the above equation, is extended over the entire domain, the evolution of the domain Ω can be handled easily by solving (10) on a regular Cartesian grid.

The algorithm starts with an initial Ω. Initializing the algorithm far from the real boundaries will require the algorithm a lot of iterations to converge. This results in a slow algorithm. Moreover, there is also the chance of the algorithm being misguided to local minima. Hence, we use a few iterations of the CG algorithm to initialize the algorithm. We define the initial potential function as

$$\phi_0(\mathbf{x}) = |u(\mathbf{x})| - T, \tag{11}$$

where $u(\mathbf{x})$ is the reconstruction given by the SIRT algorithm after a few iterations. T is an appropriately chosen threshold such that a reasonably sized region is chosen as Ω_0. From practical trials, we find that the reconstruction is not dependent on the choice of T, except for changes in the number of iterations and consequently the computation time.

Note that the expression for the velocity (8) is valid only at the boundary $d\Omega$. One approach is to use a narrow-band scheme to update the level-set in the proximity of $d\Omega$. This approach, though computationally efficient, when carefully programmed, could introduce instabilities in the evolution. Hence, we would have to re-initialize the potential function after a few iterations. Another option is to use an extension to the velocity so that the potential function is well-behaved. We use the scheme proposed in [5] since it encourages the formation of regions away from the current boundary.

$$v(\mathbf{x}) = \begin{cases} v_{\text{opt}}(\mathbf{x}) & \text{on } \Omega \cup d\Omega \\ -\mathbb{R}e\left[w^* \left(\boldsymbol{\kappa}_{D\setminus\Omega}\right)^* (\boldsymbol{\kappa}_\Omega \mathbf{f} - \mathbf{y})\right] & \text{on } D \setminus (\Omega \cup d\Omega) \end{cases} \tag{12}$$

Here, the function $w(\mathbf{x})$ is chosen as

$$w(\mathbf{x}) = \phi(\mathbf{x}) \angle (u(\mathbf{x})) \tag{13}$$

Here, $\phi(\mathbf{x})$, is the current potential function and $\angle(u(\mathbf{x}))$ is the angle of the initialization (11).

IV. RESULTS AND OBSERVATIONS

We use a Monte-Carlo simulation of the optical transport problem on a segmented MPRAGE MRI brain scan, discussed in [7], to generate the light bundles. We assumed the source-detector configuration in [7] for the simulations; 16 sources arranged in a circular pattern and 4 detectors in the center is used to obtain 64 source-detector pairs. Blob-like perturbations (see Fig. III) are inserted on the cortical region (10% of the original absorption coefficient) and the corresponding perturbed fluence measurements are obtained. Based on experimental evidence, we set the standard deviation of the noise process to be 2% of the baseline signal. Since this corresponds to a very noisy scenario, it is common practice to average N successive measurements to improve the signal to noise ratio (SNR). In our study, we typically use around 50-100 averages; the resulting temporal resolution is still sufficient for heamo-dynamic studies.

We compare the new approach to the standard approaches for DOI reconstructions: SIRT and truncated conjugate-gradients (CG) methods. Since the ground truth is difficult to obtain from a real scan, we base the comparisons entirely on simulated data. The optimal parameters (number of iterations in CG and SIRT and the parameters λ, μ and ν in the new algorithm) are determined with the knowledge of the ground truth; we determined the parameters by comparing the reconstructions to the original for a particular perturbation at the

specified noise level. This parameter set was then used for all experiments at this noise level.

In Fig 2, we demonstrate the algorithm when there are two activated regions. Note that the SIRT and CG algorithms give a uniformly smooth reconstructions; it is not possible to separate the two regions. On the other hand, the new algorithm is capable of separating the two activations, although the shapes are distorted. This indicates that the new approach can resolve near-by structures better.

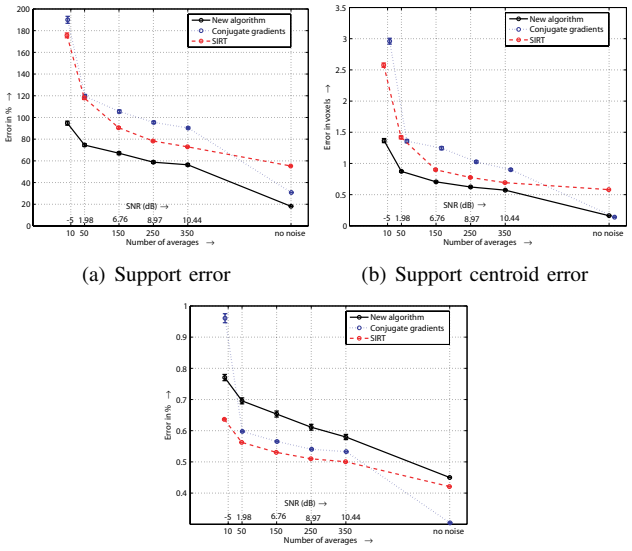
We now compare different algorithms by comparing the estimates with the ground truth f_{orig}. Here, we considered the case with only one activation. We use different metrics for comparison:

(a) Mean squared error: The mean squared error in estimation is given by $\text{MSE} = \int_{\mathbb{R}^3} \left| f^{\text{orig}}(\mathbf{x}) - f^{\text{est}}(\mathbf{x}) \right|^2 d\mathbf{x}$. While this error term is widely used and is simple, it can be misleading in our setting. Since the level-set algorithm would give piece-wise smooth reconstructions, the MSE would indicate larger errors even when the support of the reconstructions are only slightly off from the real ones.

(b) Support error (SE): Using an automatic algorithm, we classify the reconstructions into two regions (activated and background). The estimated supports of the activations are then compared to the real supports to obtain the support error. Note that the above metric accounts for the misclassified voxels in the reconstruction and ignores any amplitude changes. This metric probably makes better sense for the comparison of reconstructions of functional activations.

(c) Support centroid error (SCE): Here the distance between the centroids of the original activation and that of the detected activation is chosen as the error measure. This approach uses the classification from the earlier metric.

Using these error measures, we quantitatively compare the performance of the algorithms (new approach, CG and SIRT) as a function of the number of averages. Note that the square-root of the number of averages is inversely proportional to the SNR. The results are displayed in Fig.3. Note that the new algorithm fares better under two of the criteria (SE and SCE). However, under the mean-squares criterion (MSE), it perform worser than the others. As we have explained before, this is due to the inherent weakness of MSE to compare piece-wise constant functions.



(a) Support error

(b) Support centroid error

(c) Mean squared error

Fig. 3. Quantitative comparison of the performance of the algorithms to noise (2% of baseline signal). The errors except for SCE are displayed in percentage values of the original. Note that the new algorithm fares significantly better than the other under all criterion, except the mean-square error (MSE). We used 600 trials to generate the above plots.

V. CONCLUSIONS

We proposed a new algorithm to estimate functional activations from diffuse optical imaging. The quantitative and qualitative comparisons demonstrate the improved resolution and robustness of this algorithm over standard linear techniques.

VI. REFERENCES

[1] D. Boas, M.A. Franceschini, M. A. Dunn, and A.K. Strangman, "Noninvasive imaging of cerebral activation with diffuse optical tomography," in *In Vivo Optical Imaging of Brain Function*, R.D. Frostig, Ed., pp. 193–221. 2002.

[2] E. Gratton, V. Toronov, U. Wolf, and A. Webb, "Measurement of brain activity by near-infrared light," *J Biomed Opt.*, 2005.

[3] R. J. Gaudette, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, E. L. Miller, T. Gaudette, and D. A. Boas, "A comparison study of linear reconstruction techniques for diffuse optical tomographic imaging of absorption coefficient," *Phys. Med. Biol.*, vol. 45, pp. 1051–1070, 2000.

[4] D.L. Donoho, "For most large underdetermined systems of linear equations, the minimal l-1 norm near-solution approximates the sparsest near-solution," *preprint*, 2005.

[5] J.C. Ye, Y. Bresler, and P. Moulin, "A self-referencing level-set method for image reconstruction from sparse Fourier samples," *International Journal on Computer Vision*, vol. 50, pp. 253 –270, 2002.

[6] Y. Bresler and P.Feng, "Spectrum-blind minimum-rate sampling and reconstruction of 2-D multiband signals," in *Proc. 3rd IEEE Int. conf. on image processing, ICIP*, 1996.

[7] X. Zhang, V. Toronov, and A. G. Webb, "Simultaneous integrated diffuse optical tomography and functional magnetic resonance imaging of the human brain," *Optics express*, vol. 5513-5521, 2005.