BRAINZOOM: High Resolution Reconstruction from Multi-modal Brain Signals

Xiao Fu[*†] Kejun Huang[*†] Otilia Stretcu[*‡] Hyun Ah Song[*‡] Evangelos Papalexakis[§]
Partha Talukdar[¶] Tom Mitchell[‡] Nicholas Sidiropoulos[†] Christos Faloutsos[‡]
Barnabas Poczos[‡]

Abstract

How close can we zoom in to observe brain activity? Our understanding is limited by the resolution of imaging modalities that exhibit good spatial but poor temporal resolution, or vice-versa. In this paper, we propose BRAINZOOM, an efficient imaging algorithm that cross-leverages multi-modal brain signals. BRAINZOOM *(a) constructs high resolution* brain images from multi-modal signals, *(b) is scalable*, and *(c) is flexible* in that it can easily incorporate various priors on the brain activities, such as sparsity, low rank, or smoothness. We carefully formulate the problem to tackle nonlinearity in the measurements (via *variable splitting*) and *auto-scale* between different modal signals, and judiciously design an *inexact alternating optimization*-based algorithmic framework to handle the problem with provable convergence guarantees. Our experiments using a popular realistic brain signal simulator to generate fMRI and MEG demonstrate that high spatio-temporal resolution brain imaging is possible from these two modalities. The experiments also suggest that smoothness seems to be the best prior, among several we tried.

1 Introduction

Research in neuroimaging have resulted in the development of several techniques, viz., Electroencephalograms (EEG) [1], functional Magnetic Imaging (fMRI) [16], Magnetoencephalograms (MEG) [7], etc. These uni-modal techniques use different mechanisms to measure human brain activity. While EEG measures electrical activity of the brain, MEG measures magnetic fields generated by electric currents in the brain, and fMRI measures metabolic response due to neural activation through a blood oxygen-level dependent (BOLD) signal. Due to the diversity in the sources of these measurements, different neuroimaging techniques often have complementary strengths. For instance, while EEG and MEG have low spatial but high temporal resolution, fMRI posses just the opposite – high spatial but low temporal resolution.

Combining two or more of such complementary signals to overcome limitations posed by individual modalities has resulted in the development of several multimodal neuroimaging techniques — see [3] for a recent review. Multi-modal neuroimaging techniques have been of great interest in the brain imaging community — see [3] for a recent review. They have been used for source localization and for learning commonalities across multiple modalities, with many of these methods targeted at reducing the dimension of the brain image data, or studying specific hypotheses about relationships between imaging modalities.

However, our interest in this paper lies in reconstructing *super-resolution* images of activity across the brain – that is capturing the union of the information derived from each modality, not their intersection. While super-resolution algorithms have been applied in uni-modal settings [22], they have not been widely explored for multi-modal neuroimaging. We fill this gap by introducing BRAINZOOM, a novel method for super-resolution reconstruction of brain activity by combining different modalities of brain imaging. A snapshot of BRAINZOOM's capabilities is shown in Figure 1. Our long term goal is to complement the missing information of each modality, to ultimately construct the richest possible spatial-temporal summary of the underlying neural activity.

For the experiments in this paper, we employ BRAINZOOM to produce super-resolution brain images by combining MEG/EEG and fMRI. However, we note that BRAINZOOM is more widely applicable and may be used to create super-resolution images from other modalities as well. We make the following contributions:

[*]Equal contribution
[†]University of Minnesota
[‡]Carnegie Mellon University
[§]University of California, Riverside
[¶]Indian Institute of Science

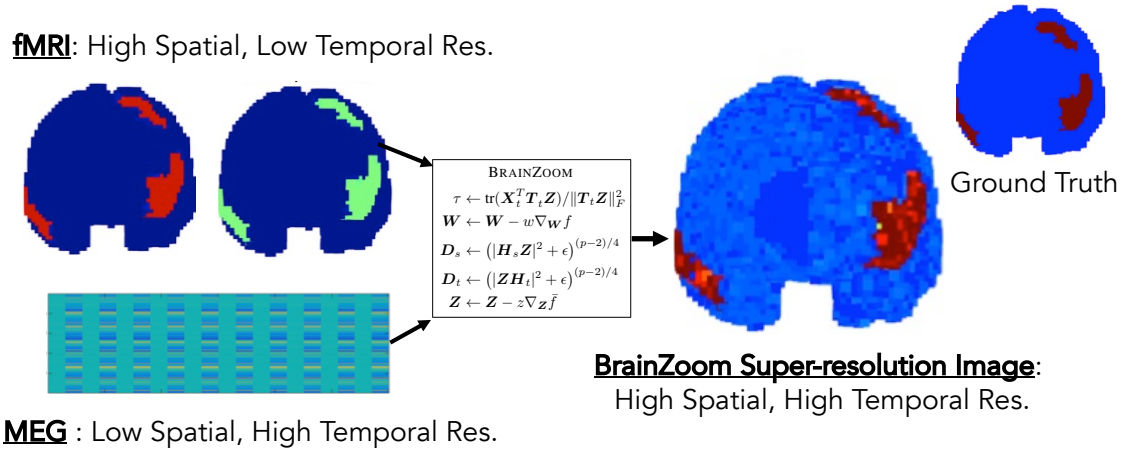


Figure 1: BRAINZOOM effectively reconstructs. BRAINZOOM effectively combines different brain measurement modalities with complementary strengths with respect to temporal and spatial resolution into a *super-resolution, whole-brain image*.

- **Novel Problem Formulation**: In Section 3 we formulate and solve the problem of generating a super-resolution brain image by combining different modalities (e.g., fMRI and MEG/EEG measurements). To the best of our knowledge, this is the first work to rigorously formulate this task as we have, as an optimization problem with this specific optimization objective, which takes into account the non-linearity in the fMRI measurements. Our formulation is also flexible in that it can easily incorporate a variety of priors that can help enhance the quality of reconstruction – so that one can seamlessly try out different hypotheses and pick out the one that gives the best performance.
- **Efficient Algorithm**: BRAINZOOM, introduced in Section 3, uses a novel modified alternating optimization method to solve the optimization problem mentioned above with high scalability. Through variable splitting, BRAINZOOM is able to handle nonlinear measurements. We also provide rigorous convergence analysis of the proposed method.
- **Evaluation**: Since the ground truth, high-resolution brain activity is not known, we cannot test the performance of our method on real brain data. Instead, in Section 4 we use simulated brain signals in order to test the ability of our method to recover the true solution. We generate synthetic brain activity using the Virtual Brain (TVB) [12] simulator. TVB is an open-source platform that creates brain activity simulations using biologically realistic assumptions, relying on well-known neuroscience models. With TVB, we can simulate synthetic brain signals with similar patterns as real brain data, as well as generate multi-modality observations such as fMRI, MEG, or EEG, at different sampling rates.
- **Reproducibility:** We will make our methods, code, and simulator scripts publicly available upon publication of the paper.

Property	*Localization*	*CCA, ICA*	*BRAINZOOM*
Hi-resolution imaging	✗	✗	✓
Scalability	✓	✓	✓
Quadratic modeling	✗	✗	✓
Convergence proof	N/A	✓	✓
Auto-scaling	✗	✓	✓
Generality	✗	✗	✓

Table 1: Comparison between BRAINZOOM and other related methods along various properties. We observe that BRAINZOOM overcomes limitations of competing methods.

2 Background and Related Work

fMRI and MEG: The exact relationship between neural activity, observed fMRI, and observed MEG signals is not fully understood. It is known that fMRI measures the ratio of oxygenated to deoxygenated hemoglobin in the blood, at a spatial resolution of approximately 1 mm, and it is widely accepted that this oxygen fluctuation is related to the convolution of local neural activity with an impulse response that lasts for 8- 10 seconds. It is also believed that fMRI corresponds more closely to local field potentials in the brain than to actual spiking [14]. In contrast, the MEG device measures magnetic fields emanating from the brain, at 1 msec temporal resolution. These magnetic fields are generated from neural currents, and are especially strong when many neurons are spatially aligned and fire synchronously. MEG signals also depend, unlike fMRI, on the direction of the neural currents generating these magnetic fields. There have been several studies of

the exact relationship between MEG and fMRI measures of neural activity. For example, Kujala et al. [11] provide evidence that fMRI correlates with source-localized MEG activity in many regions of the brain, that the exact nature of this correlation varies somewhat across the brain, and that fMRI typically correlates best with the high frequency components of MEG.

Multi-modal Methods: In the literature, there are two main classes of methods for analyzing multi-modal brain imaging data: those that aim to localize a few dominant 'point' sources of brain activity using inverse imaging or high-resolution methods; and those that target the recovery of a few dominant 'latent' components. The latter include popular multi-modal fusion techniques that rely on factor analysis tools, such as Independent Component Analysis (ICA) which aims to recover statistically independent latent sources of brain activation; and Canonical Correlation Analysis (CCA) which aims to unravel sources of co-variation in multiple 'views' of the brain stemming from the different modalities. Variants of PCA [6], ICA [5], and CCA [8] have been studied for multi-modal fusion of various brain imaging modalities. See [2,3,9,10,13,20,21] for additional background on existing localization and common component extraction methods.

The main difference between our approach and the aforementioned two broad lines of work is that we aim to reconstruct a high spatio-temporal resolution image of the entire brain – not just a few point or dominant sources / components.

Super-resolution in Image Processing: Super-resolution via fusing multi-modal data is also an active research area in image processing. Specifically, fusion of remotely sensed hyper-spectral and multi-spectral images is of great interest [15]. Both hyper- and multi-spectral images are special images whose pixels span many frequency bands and can provide spectral information of the materials contained in the pixels – which finds many applications in geoscience, mineral detection, and food safety. Hyper-spectral images have very high spectral resolution but very coarse spatial resolution, and the situation of the multi-spectral images is exactly the opposite. Much effort has been spent on fusing these two types images that captured on the sites, to create super-resolution data in both space and frequency. The task is very similar to ours, except that the transformation from the virtual super-resolution image (i.e., $\boldsymbol{Z}$) to hyper- and multi-spectral images are both linear (to be precise, sub-sampling) [19,23]. But in our problem, the transformation from $\boldsymbol{Z}$ to fMRI is nonlinear – which poses a much more challenging optimization problem.

3 Proposed Method

In this section, we first present assumptions behind our modeling of entire brain super-resolution image construction from multi-modal neuroimaging data, and then formulate it as an optimization problem. Subsequently, we discuss various priors we can add to the optimization objective. Then we propose BRAINZOOM, a highly scalable and flexible algorithmic framework that solves the above optimization problem, while provably converging to a stationary point at least sub-linearly, despite the fact that the problem is non-convex and possibly NP-hard.

3.1 Notations & Modeling Assumptions Viewing the brain as a 3D object, we can imagine its volume partitioned into many small 3D areas called voxels (equivalent to pixels in a 2D image). We are interested to know the intensity of the neural activity at each of these voxels, during a particular time frame. We start by denoting the complete brain activity as a matrix $\boldsymbol{Z}$, where each column of $\boldsymbol{Z}$ is the (vectorized) brain activity at all voxels at a particular time tick, and each row of $\boldsymbol{Z}$ corresponds to the brain activity at a particular voxel, at all recorded time ticks. However, we are not able to observe $\boldsymbol{Z}$ directly, and resort to fMRI and MEG/EEG images as inputs. The goal of this paper is to reconstruct the high-resolution brain activity matrix $\boldsymbol{Z}$ in *both* spatial and temporal domain, given the MEG and fMRI measurements.

Given the different phenomena observed by MEG and fMRI, what shall we define as the units of our latent neural activity Z? In this paper, we assume Z is a measure of the intensity of neural activity that is linearly related to the MEG or EEG observations. Let $\boldsymbol{X}_t$ denote the high temporal-resolution image given by MEG or EEG. It is related to $\boldsymbol{Z}$ through a linear operator $\boldsymbol{T}_t$ that operates on the spatial domain, possibly corrupted with noise.

$$\boldsymbol{X}_t = \boldsymbol{T}_t \boldsymbol{Z} \tag{3.1}$$

Notice that the temporal resolution of $\boldsymbol{Z}$ is the same as that of the given MEG/EEG measurements, since there is no way we can go beyond them in this dimension.

To capture the fact that fMRI is insensitive to the direction of the activity that MEG observes, we model fMRI as dependent on a linear function of the *square* of the Z activity. Although this is at best an approximation of the unknown, complex relationship between neural spiking, local field potentials, neural currents, and observed fMRI and MEG, this model does capture some basic properties, and it enables us to frame a well-posed problem of solving for high spatial and temporal resolution measure of neural activity. The success of our method in a simulator [12] of brain activity suggests that this approximation is reasonable. Let $\boldsymbol{X}_s$ be the high spatial resolution im-

age captured by fMRI. We model the dependence between $\boldsymbol{X}_s$ and Z through a linear operator T_s as shown below, again possibly corrupted with noise.

$$\boldsymbol{X}_s = \sigma|\boldsymbol{Z}|^2\boldsymbol{T}_s, \tag{3.2}$$

Here, $\sigma|\boldsymbol{Z}|^2$ is the element-wise square of $\boldsymbol{Z}$ with an unknown scaling factor σ. We note that he spatial resolution of $\boldsymbol{Z}$ is the same as that of the given fMRI images.

Reconstructing the high-resolution brain activity matrix $\boldsymbol{Z}$ in *both* the spatial and temporal domains, given the fMRI measurements $\boldsymbol{X}_s$ and MEG/EEG measurements $\boldsymbol{X}_t$, and their corresponding sensing matrices $\boldsymbol{T}_s$ and $\boldsymbol{T}_t$, is a highly ill-posed inverse problem, since the number of measurements we are given is much smaller than the number of unknowns we want to estimate. One has to impose additional constraints on the neural activity to make the problem well-defined, which will be discussed in detail in the next sections. Furthermore, considering the relationship between the underlying neural activity $\boldsymbol{Z}$ and the fMRI measurements as a *nonlinear* transformation makes the problem even harder.

3.2 Problem formulation According to the physics of multi-modal brain imaging measurements described in equations (3.1) and (3.2), the problem of infering the super-resolution description of neural activity Z can be formulated naively as the following optimization problem,

$$\min_{\boldsymbol{Z},\sigma} \|\boldsymbol{X}_t - \boldsymbol{T}_t\boldsymbol{Z}\|_F^2 + \|\boldsymbol{X}_s - \sigma|\boldsymbol{Z}|^2\boldsymbol{T}_s\|_F^2 + r(\boldsymbol{Z}), \tag{3.3}$$

where $r(\boldsymbol{Z})$ is some regularization function we impose on $\boldsymbol{Z}$ to make the estimation problem well-defined. Before we dive into different assumptions on the brain activities, let us first focus on the second term that tries to fit the fMRI data, which is a sixth-order polynomial with respect to σ and $\boldsymbol{Z}$ together, and makes the problem highly nonconvex and seemingly difficult to tackle.

In terms of σ, we notice that it is just an unknown scaling factor, and we can equivalently put it into the first term in (3.3)

$$\min_{\boldsymbol{Z},\tau} \|\boldsymbol{X}_t - \tau\boldsymbol{T}_t\boldsymbol{Z}\|_F^2 + \|\boldsymbol{X}_s - |\boldsymbol{Z}|^2\boldsymbol{T}_s\|_F^2 + r(\boldsymbol{Z}),$$

resulting the optimal solution $\boldsymbol{Z}$ to be just a scaled version of that of (3.3), which is usually inconsequential in practice. With respect to the term that involves $|\boldsymbol{Z}|^2$, we invoke the *variable splitting* trick by introducing another variable $\boldsymbol{W}$, and at the same time try to minimize the difference between $\boldsymbol{Z}$ and $\boldsymbol{W}$ in the following way

$$\begin{aligned}\min_{\boldsymbol{Z},\boldsymbol{W},\tau} \quad & \|\boldsymbol{X}_t - \tau\boldsymbol{T}_t\boldsymbol{Z}\|_F^2 + \|\boldsymbol{X}_s - (\boldsymbol{Z}*\boldsymbol{W})\boldsymbol{T}_s\|_F^2 \\ & + \mu\|\boldsymbol{Z} - \boldsymbol{W}\|_F^2 + r(\boldsymbol{Z}),\end{aligned} \tag{3.4}$$

where $*$ denotes Hadamard (element-wise) multiplication. Although the resulting formulation (3.4) is still nonconvex, the idea is to alternatingly update $\boldsymbol{Z}$, $\boldsymbol{W}$, and τ, so that the cost decreases monotonically. With careful design of the algorithm, we can even claim stronger convergence results, which will be shown.

There exist a lot of commonly used priors that can be used to reconstruct high-resolution brain signals in our formulation (3.4). In this work we consider the following possibilities:

- minimum energy: $r(\boldsymbol{Z}) = \rho\|\boldsymbol{Z}\|_F^2$;
- sparsity: $r(\boldsymbol{Z}) = \rho\|\boldsymbol{Z}\|_1$, the sum of absolute values;
- low rank: $r(\boldsymbol{Z}) = \rho\|\boldsymbol{Z}\|_*$, the sum of singular values;
- smoothness: $r(\boldsymbol{Z}) = \rho(\|\boldsymbol{H}_s\boldsymbol{Z}\|_F^2 + \|\boldsymbol{Z}\boldsymbol{H}_t\|_F^2)$, where matrices $\boldsymbol{H}_t$ and $\boldsymbol{H}_s$ have the form [4]

$$\begin{bmatrix} -1 & 2 & -1 & 0 & \dots \\ 0 & -1 & 2 & -1 & \dots \\ \vdots & & \ddots & & \\ & & -1 & 2 & -1 \end{bmatrix}, \tag{3.5}$$

  with appropriate sizes.

- total variation: $r(\boldsymbol{Z}) = \rho(\|\boldsymbol{H}_s\boldsymbol{Z}\|_{p,\epsilon}^p + \|\boldsymbol{Z}\boldsymbol{H}_t\|_{p,\epsilon}^p)$, similar to the smoothness regularization, but in this case the matrices $\boldsymbol{H}_s$ and $\boldsymbol{H}_t$ have the form

$$\begin{bmatrix} 1 & -1 & 0 & & \dots \\ 0 & 1 & -1 & & \dots \\ \vdots & & \ddots & & \\ & & & 1 & -1 \end{bmatrix}, \tag{3.6}$$

  with appropriate sizes and $\|\cdot\|_{p,\epsilon}^p$ denotes the ℓ_p quasi-norm defined as

$$\|\boldsymbol{A}\|_{p,\epsilon}^p = \sum_{i,j}(|a_{ij}|^2 + \epsilon)^{p/2}, \;\; 0 < p \leq 2.$$

  For $p = 2$, this norm reduces to the normal ℓ_2 norm that is used in the smoothness regularization; for $0 < p \leq 1$, this regularization encourages the solution to be piecewise constant, which is something we want to promote in this case.

In the rest of this section we will design an algorithm based on the latter two regularizations, which have similar forms but only differ in the definitions of the $\boldsymbol{H}$ matrices, for the following reasons:

1. Because fMRI measures convolutions of $|\boldsymbol{Z}|^2$, there is an inherent sign ambiguity in the entries of $\boldsymbol{Z}$ that are not sampled by the $\boldsymbol{T}_t$ matrix using the first three

regularizations, whereas by enforcing elements that are spatially and temporally close to each other to have similar values, this inherent sign ambiguity can be resolved. Therefore, before real validations, our conjecture is that the latter two regularizations seem to be better models.

2. Computationally, the first three regularizations can be handled by very simple *proximity operators* [17], thus it will be easy to modify the forth-coming algorithm to accommodate them. The smoothness and total-variation regularization, on the other hand, are more complicated to deal with—this is partly why we choose to use the ℓ_p quasi-norm rather than the more commonly used ℓ_1 norm to promote the effect that we want in $\boldsymbol{Z}$.

Towards this end, we formalize the problem formulation as

$$\begin{aligned}\min_{\boldsymbol{Z},\boldsymbol{W},\tau}\quad & \|\boldsymbol{X}_t-\tau\boldsymbol{T}_t\boldsymbol{Z}\|_F^2+\|\boldsymbol{X}_s-(\boldsymbol{Z}*\boldsymbol{W})\boldsymbol{T}_s\|_F^2\\ & +\mu\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2+\rho(\|\boldsymbol{H}_s\boldsymbol{Z}\|_{p,\epsilon}^p+\|\boldsymbol{Z}\boldsymbol{H}_t\|_{p,\epsilon}^p),\end{aligned}\tag{3.7}$$

and we denote the cost function in (3.7) as $f(\boldsymbol{Z},\boldsymbol{W},\tau)$.

3.3 Proposed method Based on the formulation, we propose BRAINZOOM, a highly scalable and flexible algorithmic framework for the brain super-resolution problem based on alternating optimization.

Let us start with the update of τ, which is independent of $\boldsymbol{W}$, and by fixing $\boldsymbol{Z}$, the conditionally optimal update of τ is

$$\tau\leftarrow \mathrm{tr}(\boldsymbol{X}_t^T\boldsymbol{T}_t\boldsymbol{Z})/\|\boldsymbol{T}_t\boldsymbol{Z}\|_F^2.\tag{3.8}$$

As for the update of $\boldsymbol{W}$, to avoid heavy computations, we propose to perform just one simple gradient step, where the gradient with respect to $\boldsymbol{W}$ can be calculated as

$$\begin{aligned}\nabla_{\boldsymbol{W}}f=\boldsymbol{Z}*\Big[(\boldsymbol{Z}*\boldsymbol{W})\boldsymbol{T}_s\boldsymbol{T}_s^T\Big]-\boldsymbol{Z}*(\boldsymbol{X}_s\boldsymbol{T}_s^T)\\ +\mu(\boldsymbol{W}-\boldsymbol{Z}).\end{aligned}\tag{3.9}$$

To ensure decrease of the cost function value, we need to make sure that the step-size is less than one over the Lipschitz constant of $\boldsymbol{W}$ conditioned on $\boldsymbol{Z}$. Our selection is

$$w=1\Big/\left(\lambda_{\max}(\boldsymbol{T}_s\boldsymbol{T}_s^T)\max(|\boldsymbol{Z}|^2)+\mu\right),\tag{3.10}$$

where $\lambda_{\max}(\cdot)$ computes the largest eigenvalue of the argument matrix, and in Appendix A we prove that it ensures decrease. Notice that $\lambda_{\max}(\boldsymbol{T}_s\boldsymbol{T}_s^T)$ only needs to be computed once, can then saved for subsequent iterations.

Now we turn to the update of $\boldsymbol{Z}$, and we again propose to go just a similar gradient step to avoid heavy computations. However, the regularization terms $\|\boldsymbol{H}_s\boldsymbol{Z}\|_{p,\epsilon}^p+\|\boldsymbol{Z}\boldsymbol{H}_t\|_{p,\epsilon}^p$ are non-differentiable. To compensate this, we first invoke the following property for the ℓ_p quasi-norm:

$$\|\boldsymbol{A}\|_{p,\epsilon}^p\le\sum_{i,j}\frac{p}{2}\left(|\tilde{a}_{ij}|^2+\epsilon\right)^{(p-2)/2}|a_{ij}|^2+\text{constant},$$

for all $\tilde{\boldsymbol{A}}$, and the equality holds when $\boldsymbol{A}=\tilde{\boldsymbol{A}}$. Using this property, we define $\bar{f}(\boldsymbol{Z};\boldsymbol{W},\tau)$ as

$$\begin{aligned}\bar{f}(\boldsymbol{Z};\boldsymbol{W},\tau)=&\|\boldsymbol{X}_t-\tau\boldsymbol{T}_t\boldsymbol{Z}\|_F^2+\|\boldsymbol{X}_s-(\boldsymbol{Z}*\boldsymbol{W})\boldsymbol{T}_s\|_F^2\\ &+\rho(\|\boldsymbol{D}_s*(\boldsymbol{H}_s\boldsymbol{Z})\|_F^2+\|\boldsymbol{D}_t*(\boldsymbol{Z}\boldsymbol{H}_t)\|_F^2)\\ &+\mu\|\boldsymbol{Z}-\boldsymbol{W}\|_F^2,\end{aligned}\tag{3.11}$$

where

$$\begin{aligned}\boldsymbol{D}_s&=\left(|\boldsymbol{H}_s\tilde{\boldsymbol{Z}}|^2+\epsilon\right)^{(p-2)/4},\\ \boldsymbol{D}_t&=\left(|\tilde{\boldsymbol{Z}}\boldsymbol{H}_t|^2+\epsilon\right)^{(p-2)/4},\end{aligned}\tag{3.12}$$

and $\tilde{\boldsymbol{Z}}$ is from the previous iteration. Notice that when $p=2$, i.e., in the smoothness case, the $\boldsymbol{D}$ matrices are simply the all one matrices, which makes $\bar{f}=f$. In the total variation case when $0<p\le 1$, function $\bar{f}(\boldsymbol{Z};\boldsymbol{W},\tau)$ is a smooth upperbound function for $f(\boldsymbol{Z},\boldsymbol{W},\tau)$ with respect to $\boldsymbol{Z}$ while fixing $\boldsymbol{W}$ and τ. Then we can apply a similar gradient descent step with gradient calculated as

$$\begin{aligned}\nabla_{\boldsymbol{Z}}\bar{f}=&\tau^2\boldsymbol{T}_t^T\boldsymbol{T}_t\boldsymbol{Z}-\tau\boldsymbol{T}_t^T\boldsymbol{X}_t+\mu(\boldsymbol{Z}-\boldsymbol{W})\\ &+\boldsymbol{W}*\Big[(\boldsymbol{Z}*\boldsymbol{W})\boldsymbol{T}_s\boldsymbol{T}_s^T\Big]-\boldsymbol{W}*(\boldsymbol{X}_s\boldsymbol{T}_s^T)\\ &+\boldsymbol{H}_s^T\left[\boldsymbol{D}_s^2*(\boldsymbol{H}_s\boldsymbol{Z})\right]+\left[\boldsymbol{D}_t^2*(\boldsymbol{Z}\boldsymbol{H}_t)\right]\boldsymbol{H}_t^T,\end{aligned}\tag{3.13}$$

and step size chosen as

$$\begin{aligned}z=1\Big/\Big(&\tau^2\lambda_{\max}(\boldsymbol{T}_t^T\boldsymbol{T}_t)+\lambda_{\max}(\boldsymbol{T}_s\boldsymbol{T}_s^T)\max(|\boldsymbol{W}|^2)\\ &+\mu+\rho(\max(\boldsymbol{D}_s^2)c_H^2+\max(\boldsymbol{D}_t^2)c_H^2)\Big),\end{aligned}\tag{3.14}$$

where c_H is a sharp upperbound on the maximum singular value of the $\boldsymbol{H}$ matrices

$$c_H=\begin{cases}4, & \text{for smoothness (3.5),}\\ 2, & \text{for total variation (3.6).}\end{cases}\tag{3.15}$$

Again, in Appendix A we prove that this choice of step size guarantees decrease of the cost function.

Summarizing (3.8)-(3.14), we propose BRAINZOOM as the following iterative update rule

(3.16)
$$\begin{aligned}
&\text{BRAINZOOM}\\
&\tau \leftarrow \text{tr}(\boldsymbol{X}_t^T \boldsymbol{T}_t \boldsymbol{Z})/\|\boldsymbol{T}_t \boldsymbol{Z}\|_F^2\\
&\boldsymbol{W} \leftarrow \boldsymbol{W} - w\nabla_{\boldsymbol{W}} f\\
&\boldsymbol{D}_s \leftarrow \left(|\boldsymbol{H}_s \boldsymbol{Z}|^2 + \epsilon\right)^{(p-2)/4}\\
&\boldsymbol{D}_t \leftarrow \left(|\boldsymbol{Z}\boldsymbol{H}_t|^2 + \epsilon\right)^{(p-2)/4}\\
&\boldsymbol{Z} \leftarrow \boldsymbol{Z} - z\nabla_{\boldsymbol{Z}} \bar{f}
\end{aligned}$$

3.4 Convergence Properties Although BRAINZOOM is a very simple algorithm that blends alternating optimization with gradient descent, it has very nice convergence properties, as we show here.

Theorem 1 *Let $\{(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r)})\}_r$ be the solution sequence produced by the proposed* BRAINZOOM (3.16). *We have*
1) Every limit point of $\{(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r)})\}_r$ is a stationary point of Problem (3.7).
2) The optimality gap between $\{(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r)})\}_r$ and a stationary point is at most $\mathcal{O}(1/r)$; i.e., the algorithm approaches a stationary point at least sub-linearly.

The proof of Theorem 1 is relegated to Appendix B.

Remark. The proposed BRAINZOOM is a flexible combination between alternating optimization and gradient descent. A clear advantage is that all the variations that people have been using to modify the algorithm to accommodate constraints and non-linearities, by simply putting an additional projection or proximal step. This includes the first three regularizations that we discussed in Section 3.2, namely, minimum energy, sparsity, and low rank. Moreover, if we have the prior knowledge that the underlying signal $\boldsymbol{Z}$ is non-negative, we could also add a simple projection step to zero out the negative entries. All the convergence results can be similarly applied.

4 Experiments

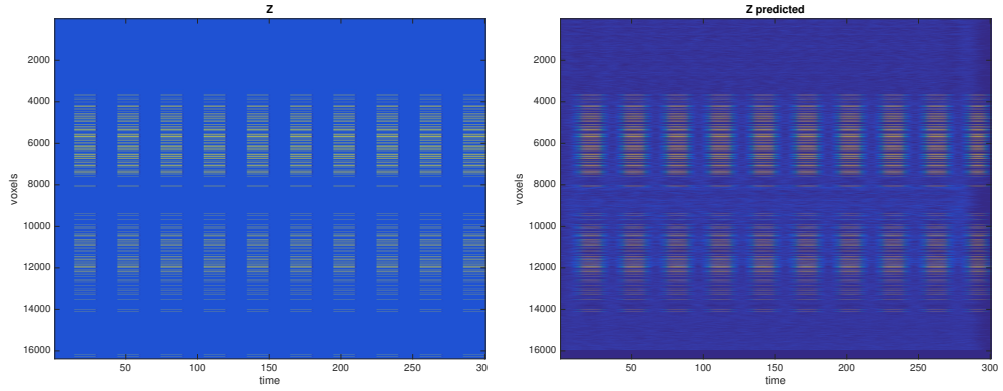
We evaluate BRAINZOOM on simulated brain data. Since in real neuroscience data we do not have access to the ground truth super-resolution brain activity, it is difficult to assess the performance our algorithm. Therefore, instead of using unattainable ground truth brain images, we generate realistic brain signals using an open source framework, the Virtual Brain (TVB) [12], which is able to simulate brain networks with biologically realistic connectivity. TVB implements several well-known neuroscience models, and can produce brain signals with similar patterns as real brain data, as well as generate multi-modal observations such as fMRI, MEG, or EEG, at different sampling rates.

In the following few lines, we describe details of the simulation. First, we select a type of stimulus, and apply it to the brain at a location and time of choice. In this experiment, we use a 3s pulse stimulus, with 3s delay. TVB records the response of the brain to this stimulus, at a very high frequency, with an integration step size of 0.0625ms. From this response we generate our $\boldsymbol{Z}$ as the temporal average of this raw data, sampled at 200ms.

We show $\boldsymbol{Z}$ in matrix form in Figure 2, where the horizontal axis represents time (we get 300 samples in time, over a total of 60s), and the vertical axis represents 16384 voxels (vectorized). Next, TVB can create multi-modality observations of this underlying signal. We sample MEG data at 200ms, and fMRI data at 1000ms. The MEG data is produced by TVB by projecting the underlying signal to the surface of the brain, at the predefined location of 248 MEG sensors. This projection is also encoded in our T_t matrix. The fMRI data is obtained by convolving the brain signal with a hemodynamic response function (hrf) of choice. Here, we chose the Gamma hrf that is given by $f(t) = (\frac{t}{\tau})^{(n-1)} * \frac{exp(-\frac{t}{\tau})}{\tau(n-1)!}$ with the default parameters suggested by TVB. We also encode this hrf in our T_s matrix. Note that TVB applies this convolution directly on underlying signal $\boldsymbol{Z}$, resulting in a simple linear transformation. Since in our method we can model a more complex relationship between fMRI and $\boldsymbol{Z}$ that TVB does not capture, and we want to test if our algorithm can recover the underlying signal under this condition, we modify generative process of the fMRI measurements to capture this non-linearity – i.e., we set the fMRI signal to be the convolution of $|Z|^2$ and the hrf described above.

Using the fMRI, MEG, T_s and T_t thus constructed, we apply our algorithm to recover the super-resolution brain $\boldsymbol{Z}$. We show $\boldsymbol{Z}$, MEG, and fMRI as space $\times$ time matrices in Figures 3 and 4, respectively. Figure 2 shows our reconstructed $\boldsymbol{Z}$, using the smoothing regularizer. One can see that the reconstructed $\boldsymbol{Z}$ matches the ground truth very well, although the recovering problem is a challenging under-determined non-linear inverse problem. This suggests that the proposed algorithm combined with the smoothness regularization is effective in fusing MEG and fMRI data.

We also map the above space$\times$time matrices back to brain manifolds at different time points and show some interesting results in Figure 5. Here, we present two fMRI images as in the right column, which are recorded at time points 38.0s and 40.0s, respectively. Before 38.0s, there was no activity in the brain (cf. the upper-left subfigure). At time point 38.0s, a stimulus is presented at three regions of the brain, and the activity spans a



(a) Ground truth. **(b)** Reconstructed.

Figure 2: Reconstructed Z matches the ground truth. Ground truth Z vs reconstructed Z.

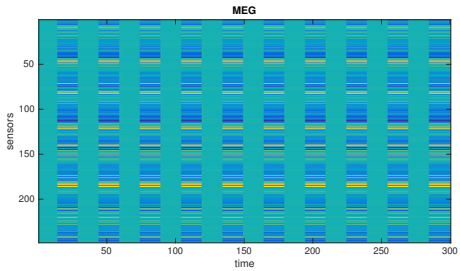


Figure 3: MEG for 248 sensors $\times$ 300 time points, across 60s of simulation.

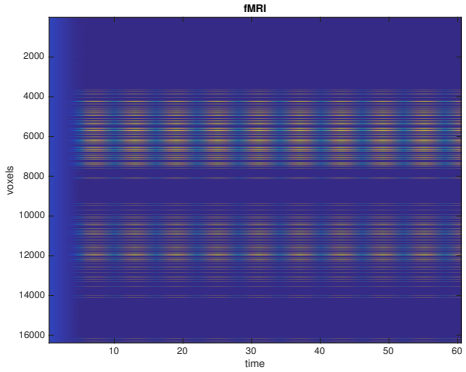


Figure 4: fMRI for 16384 voxels $\times$60 time points, across 60s of simulation.

duration starting from 38.0s across 40.0s. Our recovered Z's at 37.9s and 39.9s – when fMRI are not recorded – are presented in the middle column. One can see that the recovered Z's match the true Z's very well. This indicates that our approach correctly interpolates the time points by using the information provided by MEG, thereby achieving temporal super-resolution.

We also show the results obtained by using the minimal energy regularization in Figure 6. It can be seen that the interpolation between 39.0s and 40.0s is not as clear as we have seen in Figure 5. In addition, at the time point prior to the stimulus being present, there is an obvious artifact (bright spot) near the right edge of the estimated brain. This suggests that under the considered signal model, smoothness is a more promising regularizer.

5 Conclusions

Estimating high spatial-temporal neural activity in the brain from multiple imperfect imaging methods is a key to studying and understanding brain function. In this paper we introduce BRAINZOOM, a principled algorithmic tool to address the brain activity super-resolution problem.

Our contributions are as follows:

- **Novel Problem Formulation**: We formulate the problem of recovering a super-resolution image from multi-modal brain signals. To the best of our knowledge, this work is the first to directly address the reconstruction of a *super-resolution brain image for the whole brain*.
- **Efficient Algorithm**: We propose BRAINZOOM, an algorithmic framework that
  - handles non-linearity through variable splitting
  - learns the unknown scaling between multi-modal signals
  - incorporates a large number of regularizations
  - admits simple and efficient iterative updates
  - provably converges
- **Evaluation**: Our experimental results on the realistic simulated data matches with intuition. Both the space$\times$time plots of the estimated super-resolution matrices and the brain-manifold illustration show that our recovered space-time super-resolution brain matches with the ground-truth very well.
- **Reproducibility**: We are planning to release our source code and data upon publication. Furthermore, in order to foster and encourage reproduction and extension of our work, we describe in detail how we generate realistic data using an open source brain simulator which is widely used by neuroscientists.

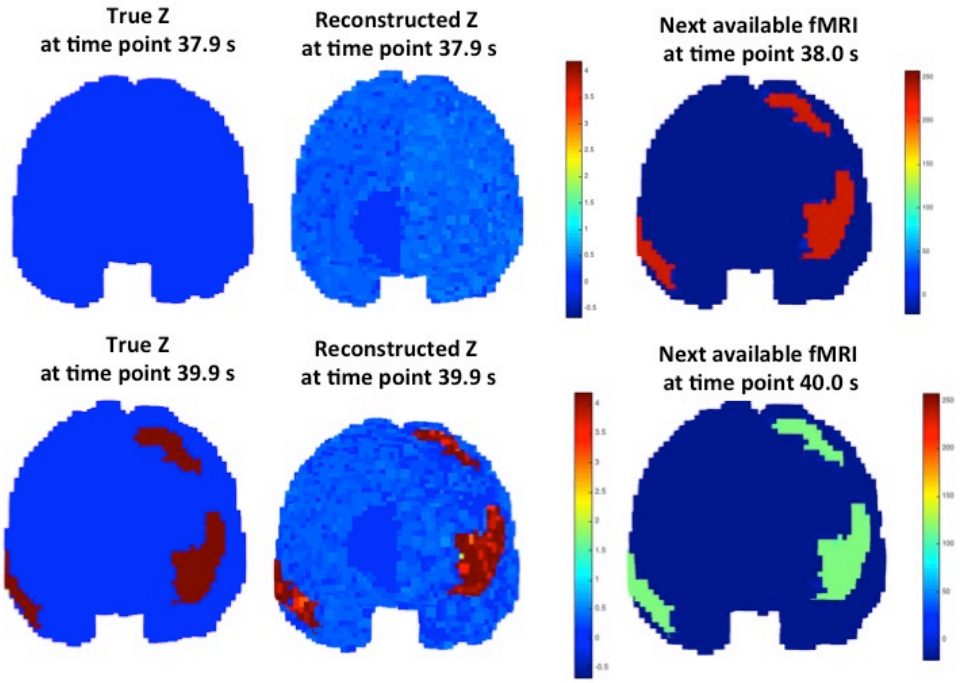


Figure 5: Smoothness regularizer recovers well. True brain activity, predicted brain activity and fMRI at two time points: when the stimulus is present (top row), and when it is absent (bottom row). Reconstructed $\boldsymbol{Z}$ is by using the smoothness regularizer and $\lambda = 1,000$.

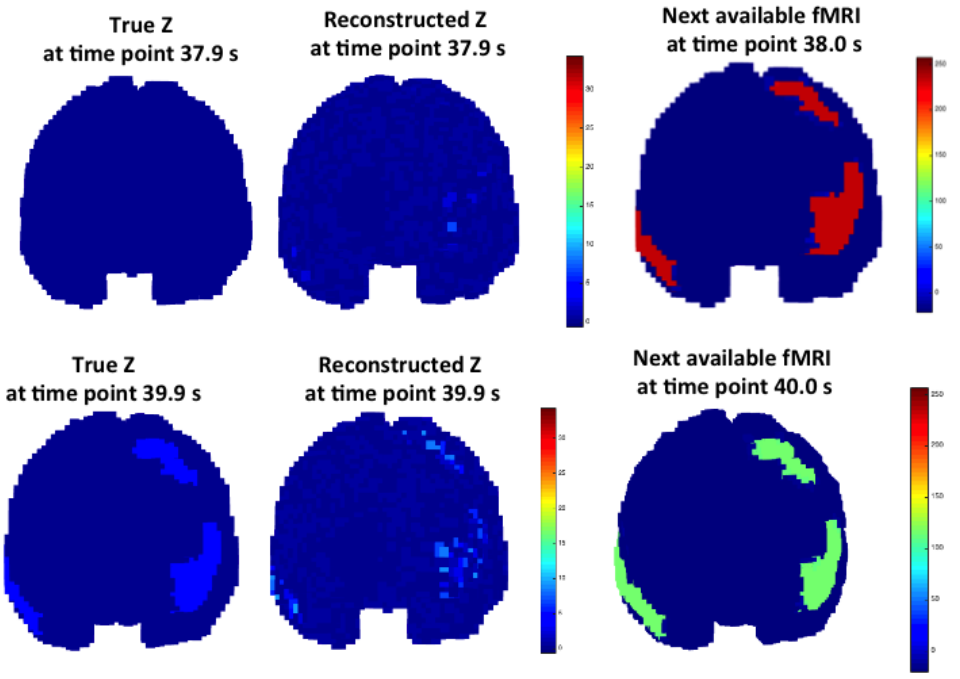


Figure 6: Minimal energy regularizer fails to recover. True brain activity, predicted brain activity and fMRI at two time points: when the stimulus is present (top row), and when it is absent (bottom row). Reconstructed $\boldsymbol{Z}$ is by using the minimal energy regularizer.

Acknowledgment

This material is based upon work supported by the National Science Foundation under Grants No. IIS-1247489, IIS-1247632.

This work is also partially supported by an IBM Faculty Award and a Google Focused Research Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, or other funding parties.

References

[1] H. Berger. Ueber das elektroenkephalogramm des menschen. *Archiv für Psychiatrie und Nervenkrankheiten*, 87(1):527–570, 1929.

[2] F. Bießmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. K. Logothetis, and K.-R. Müller. Temporal kernel cca and its application in multimodal neuronal data analysis. *Machine Learning*, 79(1-2):5–27, 2010.

[3] F. Bießmann, S. Plis, F. C. Meinecke, T. Eichele, and K.-R. Muller. Analysis of multimodal neuroimaging data. *IEEE Reviews in Biomedical Engineering*, 4:26–58, 2011.

[4] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[5] V. D. Calhoun, T. Adalı, K. A. Kiehl, R. Astur, J. J. Pekar, and G. D. Pearlson. A method for multitask fmri data fusion applied to schizophrenia. *Human brain mapping*, 27(7):598–610, 2006.

[6] A. Caprihan, G. D. Pearlson, and V. D. Calhoun. Application of principal component analysis to distinguish patients with schizophrenia from healthy controls based on fractional anisotropy measurements. *Neuroimage*, 42(2):675–682, 2008.

[7] D. Cohen et al. Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science*, 161(3843):784–786, 1968.

[8] N. M. Correa, Y.-O. Li, T. Adali, and V. D. Calhoun. Canonical correlation analysis for feature-based fusion of biomedical imaging modalities and its application to detection of associative networks in schizophrenia. *IEEE journal of selected topics in signal processing*, 2(6):998–1007, 2008.

[9] S. Dähne, F. Bießmann, F. C. Meinecke, J. Mehnert, S. Fazli, and K.-R. Müller. Integration of multivariate data streams with bandpower signals. *IEEE Transactions on Multimedia*, 15(5):1001–1013, 2013.

[10] Y. O. Halchenko, S. J. Hanson, and B. A. Pearlmutter. Multimodal integration: fmri, mri, eeg, meg. *Advanced image processing in magnetic resonance imaging*, pages 223–265, 2005.

[11] J. Kujala, G. Sudre, J. Vartiainen, M. Liljeström, T. Mitchell, and R. Salmelin. Multivariate analysis of correlation between electrophysiological and hemodynamic responses during cognitive processing. *NeuroImage*, 92:207–216, 2014.

[12] P. S. Leon, S. A. Knock, M. M. Woodman, L. Domide, J. Mersmann, A. R. McIntosh, and V. Jirsa. The virtual brain: A simulator of primate brain network dynamics. *Information-based methods for neuroimaging: analyzing structure, function and dynamics*, page 8, 2015.

[13] Z. Liu, L. Ding, and B. He. Integration of eeg/meg with mri and fmri in functional neuroimaging. *IEEE engineering in medicine and biology magazine*, 25(4):46, 2006.

[14] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fmri signal. *Nature*, 412(6843):150–157, 2001.

[15] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, et al. Hyperspectral pansharpening: a review. *IEEE Geoscience and remote sensing magazine*, 3(3):27–46, 2015.

[16] T. F. Oostendorp and A. Van Oosterom. Source parameter estimation in inhomogeneous volume conductors of arbitrary shape. *IEEE Transactions on Biomedical Engineering*, 36(3):382–391, 1989.

[17] N. Parikh and S. P. Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[18] M. Razaviyayn, M. Hong, and Z.-Q. Luo. A unified convergence analysis of block successive minimization methods for nonsmooth optimization. *SIAM Journal on Optimization*, 23(2):1126–1153, 2013.

[19] M. Simões, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot. A convex formulation for hyperspectral image superresolution via subspace-based regularization. *IEEE Transactions on Geoscience and Remote Sensing*, 53(6):3373–3388, 2015.

[20] J. Sui, T. Adali, Q. Yu, J. Chen, and V. D. Calhoun. A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of neuroscience methods*, 204(1):68–81, 2012.

[21] K. Uludağ and A. Roebroeck. General overview on the merits of multimodal neuroimaging data fusion. *NeuroImage*, 102:3–10, 2014.

[22] E. Van Reeth, I. W. Tham, C. H. Tan, and C. L. Poh. Super-resolution in magnetic resonance imaging: A review. *Concepts in Magnetic Resonance Part A*, 40(6):306–325, 2012.

[23] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *2013 IEEE ICASSP*, pages 1409–1413, 2013.

Supplementary Materials

A Step size choices.

In this appendix we show that the step sizes chosen for BRAINZOOM leads to monotonic decrease of the cost function of (3.7). The idea is to simply show that the selected step sizes are smaller than the reciprocal of the respective Liptchitz constants of the conditional functions while fixing the other variables. We show that for the step size of $\boldsymbol{Z}$, and that of $\boldsymbol{W}$ follow similar arguments.

The Liptchitz constant is the supremum of the spectral norm of the Hessian matrix of a function—for quadratic functions, it boils down to the spectral norm of Hessian. To derive it, let us first write the $\boldsymbol{Z}$-subproblem 3.11 in the vectorized form

$$\begin{aligned}\bar{f}(\boldsymbol{z};\boldsymbol{w},\tau) &= \|\boldsymbol{x}_t - \tau\bar{\boldsymbol{T}}_t\boldsymbol{z}\|^2 + \|\boldsymbol{x}_s - \bar{\boldsymbol{T}}_s\text{Diag}(\boldsymbol{w})\boldsymbol{z}\|^2 \\ &+\rho\left(\|\text{Diag}(\boldsymbol{d}_s)\bar{\boldsymbol{H}}_s\boldsymbol{z}\|^2 + \|\text{Diag}(\boldsymbol{d}_t)\bar{\boldsymbol{H}}_t\boldsymbol{z}\|^2\right) \\ &+\mu\|\boldsymbol{z}-\boldsymbol{w}\|^2,\end{aligned} \tag{A.1}$$

where the lowercase letters are vectorized versions of the matrices denoted by the uppercase letters, and

$$\bar{\boldsymbol{T}}_t = \boldsymbol{I}\otimes\boldsymbol{T}_t, \qquad \bar{\boldsymbol{T}}_s = \boldsymbol{T}_s\otimes\boldsymbol{I},$$
$$\bar{\boldsymbol{H}}_s = \boldsymbol{I}\otimes\boldsymbol{H}_s, \qquad \bar{\boldsymbol{H}}_t = \boldsymbol{H}_t\otimes\boldsymbol{I}.$$

The Hessian matrix can easily be calculated as

$$\begin{aligned}\nabla^2_{\boldsymbol{z}}\bar{f} &= \tau^2\bar{\boldsymbol{T}}_t^T\bar{\boldsymbol{T}}_t + \text{Diag}(\boldsymbol{w})\bar{\boldsymbol{T}}_s^T\bar{\boldsymbol{T}}_s\text{Diag}(\boldsymbol{w}) \\ &+\rho\bar{\boldsymbol{H}}_s^T\text{Diag}(\boldsymbol{d}_s^2)\bar{\boldsymbol{H}}_s + \rho\bar{\boldsymbol{H}}_t^T\text{Diag}(\boldsymbol{d}_t^2)\bar{\boldsymbol{H}}_t + \mu\boldsymbol{I},\end{aligned} \tag{A.2}$$

a summation of several individual matrices. An upper-bound on the spectral norm of $\nabla^2_{\boldsymbol{z}}\bar{f}$ is the sum of the individual spectral norms. Furthermore, we invoke the following properties on the matrix spectral norm:

$$\|\boldsymbol{AB}\| \le \|\boldsymbol{A}\|\|\boldsymbol{B}\|,$$
$$\|\boldsymbol{A}\otimes\boldsymbol{B}\| = \|\boldsymbol{A}\|\|\boldsymbol{B}\|.$$

Then we have that

$$\begin{aligned}\|\nabla^2_{\boldsymbol{z}}\bar{f}\| &\le \tau^2\lambda_{\max}(\boldsymbol{T}_t^T\boldsymbol{T}_t) + \max(\boldsymbol{w}^2)\lambda_{\max}(\boldsymbol{T}_t^T\boldsymbol{T}_t) \\ &+\rho\left(\max(\boldsymbol{d}_s^2)\|\boldsymbol{H}_s\|^2 + \max(\boldsymbol{d}_t^2)\|\boldsymbol{H}_t\|^2\right) + \mu.\end{aligned}$$

Finally, we show that $\|\boldsymbol{H}_s\|$ and $\|\boldsymbol{H}_t\|$ are upperbounded by c_H defined in (3.15), by taking the total variation regularization as an example, assuming its size is $(n-1)\times n$.

The definition of matrix spectral norm is

$$\|\boldsymbol{H}\| = \max_{\|\boldsymbol{u}\|=1}\|\boldsymbol{Hu}\|.$$

Assume $\hat{\boldsymbol{H}}$ is obtain by adding one more row into $\boldsymbol{H}$, then we trivially have $\|\boldsymbol{H}\| \le \|\hat{\boldsymbol{H}}\|$. Consider $\hat{\boldsymbol{H}}$ to be the following *circulant matrix*

$$\hat{\boldsymbol{H}} = \begin{bmatrix} 1 & -1 & 0 & & \dots \\ 0 & 1 & -1 & & \dots \\ \vdots & & \ddots & & \\ & & & 1 & -1 \\ -1 & 0 & \dots & 0 & 1 \end{bmatrix},$$

it can be diagonalized by the n-point discrete Fourier transform (DFT) matrix $\boldsymbol{\Phi}$

$$\hat{\boldsymbol{H}} = \boldsymbol{\Phi}\text{Diag}(\hat{\boldsymbol{h}})\boldsymbol{\Phi}^*,$$

where $\hat{\boldsymbol{h}}$ is the DFT of first row of $\hat{\boldsymbol{H}}$. Because $\boldsymbol{\Phi}$ has orthogonal columns, by rotating the elements of $\hat{\boldsymbol{h}}$ to be non-negative real, this becomes the singular value decomposition of $\hat{\boldsymbol{H}}$, and the largest absolute value of $\hat{\boldsymbol{h}}$ is the spectral norm of $\|\hat{\boldsymbol{H}}\|$. By the definition of DFT,

$$\hat{\boldsymbol{h}} = \boldsymbol{1} + [\,1\; e^{-i/n}\; e^{-i2/n}\; \dots\; e^{-i(n-1)/n}\,]^T,$$

therefore we trivially have $\max(|\hat{\boldsymbol{h}}|) \le 2$, thus

$$\|\boldsymbol{H}\| \le 2$$

Similarly, for the smoothness regularization, $\|\boldsymbol{H}\| \le 4$.

B Proof of Theorem 1

Two claims were made in Theorem 1. Here we separate them into two propositions, and prove them individually.

Proposition 1 *Let $\{(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\}_r$ be the solution sequence produced by the proposed* BRAINZOOM (3.16)*, then every limit point of $\{(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\}_r$ is a stationary point of Problem* (3.7)*.*

Proof. We prove that BRAINZOOM described in (3.16) falls into the framework of successive upper-bound minimization (BSUM) [18]. As a result, every limit point is a stationary point, according to [18, Theorem 2]. To do so, we re-write the algorithm as

$$\begin{aligned}\tau^{(r+1)} &\leftarrow \arg\min_{\tau} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau), \\ \boldsymbol{W}^{(r+1)} &\leftarrow \arg\min_{\boldsymbol{W}} u_w(\boldsymbol{W};\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}), \\ \boldsymbol{Z}^{(r+1)} &\leftarrow \arg\min_{\boldsymbol{Z}} u_z(\boldsymbol{Z};\boldsymbol{W}^{(r+1)},\boldsymbol{Z}^{(r)},\tau^{(r+1)}),\end{aligned}$$

where the $\arg\min$'s are *uniquely* defined, function u_z is an auxiliary function that satisfies that $\forall\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}$,

$$\begin{aligned}u_z(\boldsymbol{Z};\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) &\ge f(\boldsymbol{Z},\tilde{\boldsymbol{W}},\tilde{\tau}), \forall\boldsymbol{Z} \\ u_z(\tilde{\boldsymbol{Z}};\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) &= f(\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}), \\ \nabla_{\boldsymbol{Z}}u_z(\tilde{\boldsymbol{Z}};\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) &= \nabla_{\boldsymbol{Z}}f(\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}),\end{aligned}$$

and similarly for u_w.

For the update of τ, it is a scalar least-squares problem, and the minimizer is unique as long as $\|\boldsymbol{T}_t\boldsymbol{Z}\|_F^2 \neq 0$, which can be guaranteed as long as the regularization parameter ρ is not too big.

As for the auxiliary function with respect to $\boldsymbol{Z}$, we define it as

$$u_z(\boldsymbol{Z};\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) = \bar{f}(\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) + \left\langle \nabla_{\boldsymbol{Z}}\bar{f}(\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}), \boldsymbol{Z}-\tilde{\boldsymbol{Z}}\right\rangle + \frac{1}{2z}\|\boldsymbol{Z}-\tilde{\boldsymbol{Z}}\|_F^2.$$

As we have shown in Appendix A, $1/z$ is larger than the Lipschitz constant of $\bar{f}$ when fixing $\boldsymbol{W}$ and τ, therefore

$$u_z(\boldsymbol{Z};\tilde{\boldsymbol{Z}},\tilde{\boldsymbol{W}},\tilde{\tau}) \geq \bar{f}(\boldsymbol{Z};\tilde{\boldsymbol{W}},\tilde{\tau}) \geq f(\boldsymbol{Z},\tilde{\boldsymbol{W}},\tilde{\tau}).$$

It is also easy to see that the function value and gradient also coincides with that of f with respect to $\tilde{\boldsymbol{Z}}$. Furthermore, u_z is strongly convex, implying the minimizer is unique.

Similarly, and with simpler derivations, we have that u_w also satisfies the sharp upperbound requirements and the minimizer is unique. Invoking [18, Theorem 2], every limit point of BRAINZOOM is a stationary point.

Proposition 2 *In addition to Proposition 1, the optimality gap between $\{(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\}_r$ and a stationary point is at most $\mathcal{O}(1/r)$; i.e., the algorithm approaches a stationary point at least sub-linearly.*

Proof. For the τ-subproblem, because

$$\tau^{(r+1)} = \mathrm{tr}(\boldsymbol{X}_t^T\boldsymbol{T}_t\boldsymbol{Z}^{(r)})/\|\boldsymbol{T}_t\boldsymbol{Z}^{(r)}\|_F^2,$$

we have that

(B.3)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)}) - f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}) \\ &= \|\boldsymbol{X}_t - \tau^{(r)}\boldsymbol{T}_t\boldsymbol{Z}^{(r)}\|_F^2 - \|\boldsymbol{X}_t - \tau^{(r+1)}\boldsymbol{T}_t\boldsymbol{Z}^{(r)}\|_F^2 \\ &= \frac{1}{\|\boldsymbol{T}_t\boldsymbol{Z}^{(r)}\|_F^2}\left(\partial_\tau f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\right)^2 \\ &= t^{(r)}\left(\partial_\tau f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\right)^2, \end{aligned}$$

where we define $t^{(r)} = 1/\|\boldsymbol{T}_t\boldsymbol{Z}^{(r)}\|_F^2$

For the $\boldsymbol{W}$-subproblem, we have the following inequality:

(B.4)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) \\ &\leq f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}) + \frac{L_w^{(r)}}{2}\|\boldsymbol{W}^{(r)}-\boldsymbol{W}^{(r+1)}\|_F^2 \\ &\quad + \left\langle \nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}), \boldsymbol{W}^{(r+1)}-\boldsymbol{W}^{(r)}\right\rangle. \end{aligned}$$

We also notice that

(B.5)
$$\boldsymbol{W}^{(r+1)} = \arg\min_{\boldsymbol{W}} \ \frac{1}{2w^{(r)}}\|\boldsymbol{W}-\boldsymbol{W}^{(r)}\|_F^2 + \left\langle \nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}), \boldsymbol{W}-\boldsymbol{W}^{(r)}\right\rangle,$$

which means

(B.6)
$$\left\langle \nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}), \boldsymbol{W}^{(r+1)}-\boldsymbol{W}^{(r)}\right\rangle + \frac{1}{2w^{(r)}}\|\boldsymbol{W}^{(r+1)}-\boldsymbol{W}^{(r)}\|_F^2 \leq 0,$$

therefore

(B.7)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}) - f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) \\ &\geq \left(\frac{1}{2w^{(r)}} - \frac{L_w^{(r)}}{2}\right)\|\boldsymbol{W}^{(r+1)}-\boldsymbol{W}^{(r)}\|_F^2. \end{aligned}$$

On the other hand, since $\boldsymbol{W}^{(r+1)}$ is the minimizer of (B.5), by the first order optimality condition, we have

$$\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}) - \frac{1}{w^{(r)}}(\boldsymbol{W}^{(r+1)}-\boldsymbol{W}^{(r)}) = 0.$$

In sum,

(B.8)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)}) - f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) \\ &\geq \left(\frac{w^{(r)}}{2} - \frac{L_w^{(r)}}{2{w^{(r)}}^2}\right)\|\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)})\|_F^2. \end{aligned}$$

Similarly for $\boldsymbol{Z}$, we can show that

(B.9)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) - f(\boldsymbol{Z}^{(r+1)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) \\ &\geq \left(\frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2{z^{(r)}}^2}\right)\|\nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)})\|_F^2. \end{aligned}$$

Combining (B.3), (B.8), and (B.9), we obtain

(B.10)
$$\begin{aligned} &f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)}) - f(\boldsymbol{Z}^{(r+1)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)}) \\ &\geq t^{(r)}\left(\partial_\tau f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\right)^2 \\ &\quad + \left(\frac{w^{(r)}}{2} - \frac{L_w^{(r)}}{2{w^{(r)}}^2}\right)\|\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)})\|_F^2 \\ &\quad + \left(\frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2{z^{(r)}}^2}\right)\|\nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)})\|_F^2. \end{aligned}$$

To show the convergence rate, let us define

$$\begin{aligned} \phi^{(r)} = &\left(\partial_\tau f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r)})\right)^2 \\ &+ \|\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r)},\tau^{(r+1)})\|_F^2 \\ &+ \|\nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)},\boldsymbol{W}^{(r+1)},\tau^{(r+1)})\|_F^2 \end{aligned}$$

One can see that

$$\begin{aligned}
\phi^{(r)} \to 0 \qquad & \Longrightarrow \\
& \left(\partial_\tau f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r)})\right)^2 \to 0 \\
& \|\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r+1)})\|_F^2 \to 0 \\
& \|\nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r+1)}, \tau^{(r+1)})\|_F^2 \to 0
\end{aligned}$$

which implies a stationary point is attained. Let us assume that the first time $\phi^{(r)} < \varepsilon$ requires T iterations. Then, summing up (B.10) over $r = 1, ..., T$, we have

$$\begin{aligned}
& f(\boldsymbol{Z}^{(1)}, \boldsymbol{W}^{(1)}, \tau^{(1)}) - f(\boldsymbol{Z}^{(T)}, \boldsymbol{W}^{(T)}, \tau^{(T)}) \\
& \qquad \geq \sum_{r=1}^{T} t^{(r)} \left(\partial_\tau f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r)})\right)^2 \\
& + \sum_{r=1}^{T} \left(\frac{w^{(r)}}{2} - \frac{L_w^{(r)}}{2{w^{(r)}}^2}\right) \|\nabla_{\boldsymbol{W}} f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r)}, \tau^{(r+1)})\|_F^2 \\
& + \sum_{r=1}^{T} \left(\frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2{z^{(r)}}^2}\right) \|\nabla_{\boldsymbol{Z}} f(\boldsymbol{Z}^{(r)}, \boldsymbol{W}^{(r+1)}, \tau^{(r+1)})\|_F^2. \\
& \qquad \geq \sum_{r=1}^{T} c\phi^{(r)} \qquad \text{(B.11)}
\end{aligned}$$

where

$$c = \min_r \left\{ t^{(r)}, \left(\frac{w^{(r)}}{2} - \frac{L_w^{(r)}}{2{w^{(r)}}^2}\right), \left(\frac{z^{(r)}}{2} - \frac{L_z^{(r)}}{2{z^{(r)}}^2}\right) \right\}.$$

The above implies

$$\frac{f(\boldsymbol{Z}^{(1)}, \boldsymbol{W}^{(1)}, \tau^{(1)}) - f(\boldsymbol{Z}^{(T)}, \boldsymbol{W}^{(T)}, \tau^{(T)})}{T} \geq c\phi^{(r)},$$

and so

$$\phi^{(r)} \leq \frac{1}{T}\left(\frac{f(\boldsymbol{Z}^{(1)}, \boldsymbol{W}^{(1)}, \tau^{(1)}) - f(\boldsymbol{Z}^{\star}, \boldsymbol{W}^{\star}, \tau^{\star})}{c}\right),$$

where $\boldsymbol{Z}^{\star}, \boldsymbol{W}^{\star}, \tau^{\star}$ denote a global optimal solution of Problem (3.7). This completes the proof.