SIAM J. IMAGING SCIENCES
Vol. 7, No. 1, pp. 528–557

Intensity Nonuniformity Correction for Brain MR Images with Known Voxel Classes*

Yunho Kim† and Hemant D. Tagare‡

Abstract. Intensity nonuniformity in magnetic resonance (MR) images, represented by a smooth and slowly varying function, is a typical artifact that is a nuisance for many image processing methods. To eliminate the artifact, we have to estimate the nonuniformity as a smooth and slowly varying function and factor it out from the given data. We reformulate the problem as a problem of finding a unique smooth function in a particular set of piecewise smooth functions and propose a variational method for finding it. We deliver the main idea using a simple one-dimensional example first and provide a thorough analysis of the problem in a three-phase scenario in three dimensions whose application can be found in the brain MR images. Experiments with synthetic and real MR images and a comparison with a state-of-the-art method, N3, show our algorithm's satisfactory performance in estimating the nonuniformity with and without noise. An automated procedure is also proposed for practical use.

Key words. intensity nonuniformity, magnetic resonance imaging, image segmentation

AMS subject classifications. 65K10, 68U10, 49J40

DOI. 10.1137/130924688

1. Introduction.

Magnetic resonance (MR) images often contain an artifact called the "intensity nonuniformity artifact" [2]. Possible causes of this artifact are RF coil inhomogeneity and gradient-driven eddy currents. This artifact causes MR image intensities to be spatially modulated by a function. The standard model for the nonuniformity artifact is

$$h = g \cdot f + \eta, \tag{1.1}$$

where h is the observed image and g, f, and η are the modulating function, the true image, and noise, respectively. In many applications, estimating and compensating for the modulation function g is necessary before further processing the image. In the literature, the modulating function g is sometimes called the *bias field* or the *nonuniformity*.

As (1.1) stands, the nonuniformity function g cannot be estimated uniquely from h because, a priori, h cannot be uniquely factored into two functions. Additional assumptions are required. The usual assumptions are that g is extremely smooth while f is piecewise constant. Thus, factoring h means associating smooth changes of h with g, and the discontinuities of h with f. These assumptions eliminate most of the ambiguity in factoring h, but not all of it. It is still impossible to distinguish between factorizations of h that differ by multiplicative constants, i.e., between the factorization $h = g \cdot f$, and the factorization $h = (ag) \cdot (f/a)$ for

*Received by the editors June 12, 2013; accepted for publication (in revised form) December 17, 2013; published electronically March 13, 2014.
http://www.siam.org/journals/siims/7-1/92468.html
†Department of Diagnostic Radiology, Yale University, New Haven, CT 06520 (yunho.kim@yale.edu).
‡Department of Diagnostic Radiology, Biomedical Engineering, Yale University, New Haven, CT 06520 (hemant.tagare@yale.edu).

$a > 0$. We live with this ambiguity and take "estimating g from h" to mean calculating any function that is proportional to g, with a positive constant of proportionality.

There have been many previous attempts at estimating g from h; we refer the reader to a comprehensive survey [2] and the references therein. Among the nonuniformity estimation methods, the N3 method (nonparametric nonuniform intensity normalization) [14] is popular. It begins with the following observation: Without noise, applying the logarithmic function to the noiseless version of (1.1) gives

$$\log(h) = \log(g) + \log(f). \tag{1.2}$$

Thus the problem of finding the multiplicative function g from (1.1) becomes the problem of finding the additive function $\log(g)$ from (1.2), which is more tractable. Assuming that $\log(g)$ and $\log(f)$ are independent or uncorrelated random variables and that H, G, F are probability density functions of $\log(h), \log(g), \log(f)$, respectively, the problem becomes estimating the probability density G of $\log(g)$. The relation (1.2) implies that $H = G * F$, which shows that the effect of the nonuniformity distribution G is to blur the probability density F. The N3 algorithm first estimates F by sharpening the observed H and with this estimate $\tilde{F}$, a Gaussian function G is found that produces the best fit to the estimate $\tilde{F}$ given H via deconvolution. Then producing $\log(g)$ involves another smoothing process using B splines. It is reported in [14] that the quality of the result produced by N3 is senitive to the smoothing process. In addition, N3 has many parameters that have to be tweaked for a good estimate. For example, it is observed in [1] that the parameters used in the past for 1.5 T scanners needed modification for 3 T scanners. Finally, a theoretical limitation of N3 is that its convergence has not been mathematically analyzed and is poorly understood.

There are more recent algorithms that use variational approaches for nonuniformity correction [9, 12]. The authors of [9] use the fact that images from a body coil have low signal-to-noise ratio but good spatial homogeneity, while images from surface coils have strong signal response near the coils but the image intensity rapidly diminishes away from the coils. Making use of data obtained from a body coil and surface coils, the authors propose a discrete model:

$$\min_{\mathbf{b},\mathbf{f}} \lambda_B \|\mathbf{y}_B - \mathbf{f}\|^2 + \lambda_S \|\mathbf{y}_S - \mathbf{b} \circ \mathbf{f}\|^2 + \alpha \|\mathbf{L}\mathbf{b}\|^2 + \gamma \|\mathbf{D}\mathbf{f}\|_p^p,$$

where B and S stand for body and surface, respectively, $\mathbf{L}$ is the Laplace operator, and $\mathbf{D}$ is the gradient operator. The bias field $\mathbf{b}$ and the underlying structure $\mathbf{f}$ are estimated by the above minimization problem, which is extended to situations where multiple surface coils and multiple pulse sequences are used. In contrast to the above model, the authors of [12] propose a continuous model:

$$\min_{u,\sigma} J(u,\sigma) = \frac{1}{2}\int_\Omega |\sigma u - \tilde{u}|^2 dx + \frac{\nu}{2}\int_\Omega |\nabla^2 \sigma|^2 dx + \frac{\kappa}{2}\int_\Omega u^2 dx + \mu \int_\Omega \phi_\epsilon(|Du|) dx,$$

and analyze the minimization problem in the primal-dual setting. Here σ is the bias field and u is the image without nonuniformity. Both of these variational approaches model the bias field in the Sobolev space H^2 and the underlying structure either in the Sobolev space $W^{1,p}$ or in the BV space.

Unlike the above approaches, our algorithm is designed for cases where some knowledge is available of anatomy contained in the MR image. This is often true of large structural studies of the brain for various disorders or diseases with a structural component (e.g., images such as those in the publicly available PPMI database). In these studies, once the skull is stripped from the brain images, the remaining pixels belong to three classes: cerebro-spinal fluid (CSF), gray matter (GM), and white matter (WM). Moreover, for a given field strength, say 1.5 T, the relative brightness of the three pixel classes is approximately known. For example, at 1.5 T WM is roughly twice as bright as CSF. As we show below, this prior knowledge can be exploited to accurately estimate the nonuniformity function in a variational formulation of the problem. Our goal is to solve the intensity nonuniformity problem using this idea in a mathematically sound formulation, taking care to state the underlying assumptions clearly and to prove convergence. Our formulation turns out to be closely related to the classical problem of variational image segmentation. So we draw upon variational segmentation theory to design our algorithm, and to prove its convergence.

The outline of this paper is as follows. In section 2, we informally discuss the key idea behind our algorithm using a rather simple one-dimensional (1-D) example. In section 3, we present the mathematical framework that we use to pose the problem precisely. In section 4, we pursue a detailed analysis of our proposed problem for a three-phase scenario with an application to the brain MR imaging in mind. In section 5, we extend our three-phase framework to a general K-phase case, whose analysis is not further repeated. In section 6, we provide numerical schemes and some explanation about choosing algorithm parameters. In section 7, we further propose an adaptive procedure which turns out to be useful in practice when prior knowledge is not accurate. In section 8, we report the performance of our algorithm in estimating the intensity nonuniformity in synthetic MR images with and without noise. We also explain the need for an automated procedure in practice and show how the adaptive procedure from section 7 works. Then, we provide results of using our algorithms with real MR images and close the section by comparing our results with the state-of-the-art N3 method.

2. Nonuniformity correction using prior knowledge.

2.1. A simple 1-D example. To informally explain the main idea, we begin with a highly simplified example, where the true image f and the nonuniformity function g are 1-D functions on an interval $\mathcal{I}$ of the real line, and the observed image h is noise-free. This is illustrated in Figure 1. We further assume for simplicity that the underlying image f is piecewise constant, as shown in Figure 1(a), so that

$$f = \mu_1 1_{\Sigma_1} + \mu_2 1_{\Sigma_2},$$

where 1_{Σ_1} and 1_{Σ_1} are indicator functions of two disjoint sets that partition $\mathcal{I}$, in which f takes values μ_1 and μ_2, respectively. We consider these sets to be two disjoint intervals only for illustrative purposes. The explanation below requires only that the sets partition $\mathcal{I}$. We note that $\mu_1, \mu_2, \Sigma_1, \Sigma_2$ are all unknown.

Let g be the nonuniformity function as shown in Figure 1(b), and let $h = f \cdot g$ be the observed image as in Figure 1(c). The task is to recover from h a function proportional to g without further knowledge of f. This can be done as follows: Let $\rho = \mu_1/\mu_2$ be the ratio of the μ's, and assume that ρ is known. Set $h_1 = h\sqrt{\rho}$ and $h_2 = h/\sqrt{\rho}$. The functions h_1

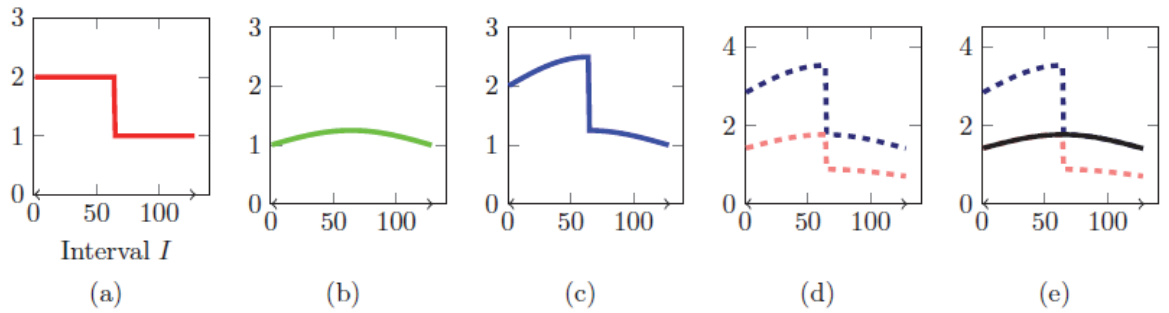


Figure 1. (a) *Underlying signal f defined on an interval I, which is piecewise constant taking two values;* (b) *nonuniformity function g;* (c) *the function $h = g \cdot f$;* (d) *plot of $h\sqrt{\rho}$ in dark blue dashed line and $h/\sqrt{\rho}$ in light red dashed line;* (e) *the black solid curve is the information we can retrieve about the nonuniformity function g, which is $\sqrt{\mu_1\mu_2}g$.*

and h_2 are drawn in Figure 1(d) with dashed lines. Notice that if we combine a piece of h_1 with a piece of h_2, then we can recover a function proportional to g. This is illustrated in Figure 1(e), where the combined function is drawn with a solid line. To be more precise, the functions h_1, h_2 are given by

$$h_1 = h\sqrt{\rho} = \sqrt{\frac{\mu_1}{\mu_2}}g \times (\mu_1 1_{\Sigma_1} + \mu_2 1_{\Sigma_2}) = \sqrt{\frac{\mu_1}{\mu 2}}g\mu_1 1_{\Sigma_1} + \sqrt{\mu_1\mu_2}g1_{\Sigma_2},$$
$$h_2 = h/\sqrt{\rho} = \sqrt{\frac{\mu_2}{\mu_1}}g \times (\mu_1 1_{\Sigma_1} + \mu_2 1_{\Sigma_2}) = \sqrt{\mu_1\mu_2}g1_{\Sigma_1} + \sqrt{\frac{\mu_2}{\mu_1}}\mu_2 g1_{\Sigma_2}.$$

Since $1_{\Sigma_1}1_{\Sigma_1} = 1_{\Sigma_2}1_{\Sigma_2} = 1$ and $1_{\Sigma_1}1_{\Sigma_2} = 1_{\Sigma_2}1_{\Sigma_1} = 0$, we have $h_1 1_{\Sigma_2} + h_2 1_{\Sigma_1} = g\sqrt{\mu_1\mu_2}$, which is proportional to g.

The above intuition suggests a possible algorithm: Calculate h_1 and h_2 from h by multiplying and dividing by $\sqrt{\rho}$. Estimate a partition Σ_1, Σ_2 of $\mathcal{I}$ such that $h_1 1_{\Sigma_2} + h_2 1_{\Sigma_1}$ gives the smoothest possible function. Note that the only way to do this in Figure 1 is to partition $\mathcal{I}$ exactly at the discontinuity of h. Then, the function $h_1 1_{\Sigma_2} + h_2 1_{\Sigma_1}$ is an estimate of a function proportional to g. Basically, this is our algorithm. An important technical part of the problem is to estimate Σ_1, Σ_2 by maximizing the regularity of $h_1 1_{\Sigma_2} + h_2 1_{\Sigma_1}$. This subproblem is closely related to segmentation.

The idea discussed above can be easily extended to the case where f is piecewise constant and takes $K \geq 2$ different values (the case above has $K = 2$). But formulating the problem in this generality gives equations that are somewhat opaque. To make the presentation simple, we formulate the problem and present its analysis for $K = 3$ in section 4. This case is important because one of our main applications is correcting for intensity inhomogeneities in skull-stripped brain MR images, where K equals 3.

All of the definitions, theorems, and proofs for the $K = 3$ case carry over to the general $K \geq 2$ case without any difficulties.

3. Mathematical tools.

We now turn to discussing some of the mathematical background we need to formulate the problem for the $K = 3$ case.

3.1. Function spaces. We consider a bounded open set $\Omega \in \mathbb{R}^n$. For the experiments in section 6, $n = 3$ is used. The main function spaces that we rely on are the bounded variation (BV) space and the Sobolev spaces.

Definition 3.1. *A function $u \in L^1(\Omega)$ is of bounded variation if*

$$\sup \left\{ \int_\Omega u(x) \operatorname{div}(\varphi(x)) dx : \varphi \in C_c^1(\Omega; \mathbb{R}^n),\ |\varphi(x)| \leq 1 \right\} < \infty.$$

We denote by $BV(\Omega)$ the set of all the L^1 functions of bounded variation.

It is important to note that $BV(\Omega)$ imposes enough regularity for its members to be almost differentiable and yet allows discontinuities to some extent, which can be seen from the theorem below. This is the main reason for using $BV(\Omega)$ in image processing.

Theorem 3.2. *A function $u \in BV(\Omega)$ has a distributional derivative Du, an n-dimensional Radon measure on Ω, whose total variation is*

$$|Du|(\Omega) = \sup \left\{ \int_\Omega u(x) \operatorname{div}(\varphi(x)) dx : \varphi \in C_c^1(\Omega; \mathbb{R}^n),\ |\varphi(x)| \leq 1 \right\}.$$

We will denote the total variation of the Radon measure Du associated with $u \in BV(\Omega)$ by either of the following:

$$|Du|(\Omega), \quad \int_\Omega |\nabla u|.$$

If u is differentiable, then Du is simply the gradient ∇u of u and the total variation is nothing but the L^1 norm of ∇u, i.e.,

$$|Du|(\Omega) = \int_\Omega |\nabla u| = \int_\Omega |\nabla u(x)| dx.$$

Then $BV(\Omega)$ becomes a Banach space with the norm

$$\|u\|_{BV(\Omega)} = \|u\|_{L^1(\Omega)} + \int_\Omega |\nabla u|.$$

However, a better topology on $BV(\Omega)$ to work with is the weak-* topology. In particular, a set E is of finite perimeter in Ω if the characteristic function 1_E belongs to $BV(\Omega)$ and its boundary measure, the perimeter of E, will be denoted by

$$\mathrm{P}(E; \Omega) = |D1_E|(\Omega) = \int_\Omega |\nabla 1_E|.$$

The lower-semicontinuity property, the approximation property of $u \in BV(\Omega)$ by smooth functions, the compactness property of a bounded sequence in $BV(\Omega)$, and the coarea formula are particularly useful in analyzing functionals defined on $BV(\Omega)$. For more details about the properties of $BV(\Omega)$, we refer the reader to [8] and [10].

Another function space that we use is the Sobolev space $H^k(\Omega)$ defined as follows.

Definition 3.3. *Let $k \in \mathbb{N}$. Then $H^k(\Omega)$ is the set of functions that have weak derivatives*

$$\nabla^\alpha u = \frac{\partial^{|\alpha|} u}{\partial_{x_1}^{\alpha_1} \cdots \partial_{x_k}^{\alpha_k}},$$

where

$$\alpha = (\alpha_1, \ldots, \alpha_k),\ \alpha_i \geq 0,\ i = 1, \ldots, k,\ \text{and } |\alpha| = \sum_{i=1}^{k} \alpha_i \leq k,$$

with the norm

$$\|u\|_{H^k(\Omega)} = \left(\sum_{|\alpha| \leq k} \int_\Omega |\nabla^\alpha u(x)|^2 dx \right)^{\frac{1}{2}}.$$

What is important is that $H^k(\Omega)$ is a Hilbert space for all $k \in \mathbb{N}$. We will use elements of $H^1(\Omega)$ to model the nonuniformity function g in the following sections. We refer the reader to any partial differential equations textbook for more details about the Sobolev spaces (e.g., [7]).

3.2. The Mumford–Shah model. Besides function spaces, we also discuss briefly the MS (Mumford–Shah) model, from which segmentation algorithms are commonly derived. The MS model is a minimization problem of the form

$$\min_{u,K} \mathcal{F}(u, K) = \int_{\Omega \setminus K} |\nabla u(x)|^2 dx + \alpha \int_\Omega |u(x) - f(x)|^2 dx + \beta \mathcal{H}^1(K), \tag{3.1}$$

where $\alpha, \beta > 0$ and Ω is a bounded open subset of $\mathbb{R}^2$ and $\mathcal{H}^1$ is the 1-D Hausdorff measure and K is a 1-D rectifiable curve. This model has been extensively analyzed, analytically as well as numerically. One of the references that we would like to point out is the CV (Chan–Vese) model [6], where the authors apply the MS model to the two-phase segmentation problem via a level set function ϕ to represent the region of interest. More precisely, if we consider (3.1) over a particular set containing only binary functions

$$c_1 1_\Sigma + c_2 1_{\Omega \setminus \Sigma},$$

then (3.1) becomes

$$\min_{c_1, c_2 \in \mathbb{R}, \Sigma \subset \Omega} \alpha \int_\Sigma |f(x) - c_1|^2 dx + \alpha \int_{\Omega \setminus \Sigma} |f(x) - c_2|^2 + \beta \int_\Omega |\nabla 1_\Sigma|. \tag{3.2}$$

The authors of [6], indeed, propose to solve

$$\min_{c_1, c_2, \phi} \left\{ \int_\Omega |\nabla H(\phi)| + \lambda \int_\Omega [(f(x) - c_1)^2 H(\phi(x)) + (f(x) - c_2)^2 (1 - H(\phi(x)))] dx \right\}, \tag{3.3}$$

where H is the Heaviside function. Note that (3.2) and (3.3) are equivalent with $\Sigma = \{\phi > 0\}$. For numerical experiments, a regularized Heaviside function H_ϵ is used in place of H. Later, Chan, Esedoglu, and Nikolova [5] proposed a convex formulation for (3.3) to solve (3.2) with c_1, c_2 fixed, i.e.,

$$\min_{0 \leq \phi \leq 1} \left\{ \int_\Omega |\nabla \phi| + \lambda \int_\Omega [(f(x) - c_1)^2 \phi(x) + (f(x) - c_2)^2 (1 - \phi(x))] dx \right\}. \tag{3.4}$$

To overcome some limitations of (3.4), the author of [13] used a strictly convex formulation to solve (3.2) with c_1, c_2 fixed, which gives segmentation with a faster algorithm. Generalizations to $\mathbb{R}^n$ of the MS and the CV models are straightforward and we use these techniques in our algorithm and refer the reader to the references mentioned above for further information.

4. The problem description.

4.1. Generalizing the idea of section 2.

With the preliminaries out of the way, we turn to taking the intuition of section 2 and converting it into a well-defined problem. As mentioned in section 2, we do this for the case where f takes $K = 3$ values, μ_1, μ_2, and μ_3. First, a straightforward generalization of the idea of section 2 is presented.

Theorem 4.1. *Let $\Omega \subset \mathbb{R}^N$ be open and connected, $N \geq 1$. Suppose that a given function h on Ω is a product of two unknown functions f and g through $h = f \cdot g$. Suppose g is positive and continuous in Ω and f is positive and piecewise constant with $K = 3$, i.e.,*

$$f = \mu_1 1_{\Sigma_1} + \mu_2 1_{\Sigma_2} + \mu_3 1_{\Sigma_3},$$

where $\mu_1 > \mu_2 > \mu_3 > 0$ and the partition $\Omega = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ is unknown. Further, suppose that the two ratios

$$\rho_1 = \frac{\mu_1}{\mu_2}, \ \rho_2 = \frac{\mu_2}{\mu_3}$$

are known. Then, there exist $\alpha_1, \alpha_2, \alpha_3$ in $\mathbb{R}$ depending only on ρ_1, ρ_2 such that the set X defined by

$$\Big\{u : \Omega \to \mathbb{R} : u(x) \in \{\alpha_1 h(x), \alpha_2 h(x), \alpha_3 h(x)\}\Big\},$$

contains a unique continuous function $\tilde{g}$, which is a positive constant multiple of the unknown function g. Moreover, $h/\tilde{g}$ is a constant positive multiple of the unknown function f.

Proof. We note that ρ_1, ρ_2 determine all the ratios

$$\frac{\mu_j}{\mu_i} = \prod_{k=j}^{i-1} \rho_k \quad \text{for} \quad 1 \leq j < i \leq 3.$$

Define each α_i, $i = 1, 2, 3$, by

$$\alpha_i = \left(\prod_{j \neq i} \frac{\mu_j}{\mu_i}\right)^{\frac{1}{3}} = \begin{cases} \frac{1}{\sqrt[3]{\rho_1^2 \rho_2}} & \text{if } i = 1, \\ \sqrt[3]{\rho_2/\rho_3} & \text{if } i = 2 \\ \sqrt[3]{\rho_1 \rho_2^2} & \text{if } i = 3. \end{cases}$$

Note that these α_i's are not zero and depend only on ρ_1, ρ_2. Then, for any i, j,

$$\alpha_i \mu_i = \sqrt[3]{\mu_1 \mu_2 \mu_3} = \alpha_j \mu_j,$$

which implies that $0 < \alpha_1 < \alpha_2 < \alpha_3$.

If we multiply the given h by α_i, then

$$\begin{aligned}\alpha_i h &= \alpha_i g \cdot (\mu_1 1_{\Sigma_1} + \mu_2 1_{\Sigma_2} + \mu_3 1_{\Sigma_3}) \\ &= \left(\sum_{j\neq i} \alpha_i \mu_j g 1_{\Sigma_j}\right) + \alpha_i \mu_i g 1_{\Sigma_i} \\ &= \left(\sum_{j\neq i} \alpha_i \mu_j g 1_{\Sigma_j}\right) + (g\sqrt[3]{\mu_1\mu_2\mu_3}) 1_{\Sigma_i}.\end{aligned}$$

This means that

$$\alpha_1 h 1_{\Sigma_1} + \alpha_2 h 1_{\Sigma_2} + \alpha_3 h 1_{\Sigma_3} = g\sqrt[3]{\mu_1\mu_2\mu_3}.$$

Therefore, $\tilde{g} = g\sqrt[3]{\mu_1\mu_2\mu_3}$ is continuous and is contained in the set X, where

$$X = \Big\{u : \Omega \to \mathbb{R} : u(x) \in \{\alpha_1 h(x), \alpha_2 h(x), \alpha_3 h(x)\}\Big\}.$$

Next, suppose $\bar{g} \in X$ is also a continuous function. Then, by the definition of X, there exists a partition $\Omega = \bar{\Sigma}_1 \cup \bar{\Sigma}_2 \cup \bar{\Sigma}_3$ such that

$$\bar{g} = \alpha_1 h 1_{\bar{\Sigma}_1} + \alpha_2 h 1_{\bar{\Sigma}_2} + \alpha_3 h 1_{\bar{\Sigma}_3}.$$

Note that

$$G := \frac{\tilde{g}}{\bar{g}} = \frac{\sum_{i=1}^3 \alpha_i 1_{\Sigma_i}}{\sum_{i=1}^3 \alpha_i 1_{\bar{\Sigma}_i}} = \sum_{1\leq i,j\leq 3} \frac{\alpha_i}{\alpha_j} 1_{\Sigma_i \cap \bar{\Sigma}_j}$$

must be continuous. Since G is piecewise constant, G must be a constant function. Suppose that $G \equiv t \neq 1$. It is easy to see that for $i = 1, 2, 3$,

$$\Sigma_i \cap \bar{\Sigma}_i = \emptyset.$$

Moreover, for any $x_1 \in \Sigma_1$ and $x_2 \in \bar{\Sigma}_1$, there exists $i, j > 1$ such that $x_1 \in \bar{\Sigma}_i$ and $x_2 \in \Sigma_j$. This implies that

$$G(x_1) = \frac{\alpha_1}{\alpha_i} < 1 < \frac{\alpha_j}{\alpha_1} = G(x_2),$$

which is a contradiction because $G \equiv t \neq 1$. Hence, $G \equiv 1$ and $\tilde{g}$ is the unique continuous function in X. Note that

$$\frac{h}{\tilde{g}} = \frac{f \cdot g}{g\sqrt[3]{\mu_1\mu_2\mu_3}} = \frac{1}{\sqrt[3]{\mu_1\mu_2\mu_3}} f$$

is a constant multiple of the unknown function f. ■

This theorem captures the intuition of section 2 precisely. It shows that up to a constant scaling factor, the function g can be found in the set

$$X = \Big\{u : \Omega \to \mathbb{R} : u(x) \in \{\alpha_1 h(x), \alpha_2 h(x), \alpha_3 h(x)\}\Big\},$$

which is the only continuous function in X; all other functions in X have some discontinuity.

For simplicity, here and in what follows, we will reserve the symbol ψ for "g up to a multiplicative factor," which is the unique continuous function in the set X we want to find.

4.2. Finding ψ. How can we use Theorem 4.1 when there is noise in the data? Naively, we might try to find ψ by minimizing the L^2 norm of

$$\psi - (\alpha_1 h 1_{\Sigma_1} + \alpha_2 h 1_{\Sigma_2} + \alpha_3 h 1_{\Sigma_3})$$

with some regularization penalty for ψ. But this idea does not correctly handle the noise variance. According to (1.1), the noise in h has equal variance at all points of Ω. Multiplying h with different α's in Σ_1, Σ_2, and Σ_3 will result in unequal noise variances in Σ_1, Σ_2, and Σ_3. Simply minimizing the L^2 norm of $\psi - (\alpha_1 h 1_{\Sigma_1} + \alpha_2 h 1_{\Sigma_2} + \alpha_3 h 1_{\Sigma_3})$ does not account for differing noise variances in the three regions and hence is not appropriate. But it is straightforward to see that $(\psi/\alpha_1 - h)1_{\Sigma_1} + (\psi/\alpha_2 - h)1_{\Sigma_2} + (\psi/\alpha_3 - h)1_{\Sigma_3}$ retains equality of noise variance since the data, h, are not being multiplied by any α_i's. Thus, finding ψ by minimizing the L^2 norm of $(\psi/\alpha_1 - h)1_{\Sigma_1} + (\psi/\alpha_2 - h)1_{\Sigma_2} + (\psi/\alpha_3 - h)1_{\Sigma_3}$ is the appropriate use of Theorem 4.1 when the data are noisy.

What remains is to construct an algorithm for finding ψ. To this end, we represent the partition of Ω into three disjoint sets by two functions $0 \le \phi_1, \phi_2 \le 1$, so that $1_{\Sigma_1}, 1_{\Sigma_2}$, and 1_{Σ_3} are closely approximated by $\phi_2\phi_1$, $\phi_2(1-\phi_1)$, and $(1-\phi_2)$, respectively. The regularity of ψ is measured by a functional $\mathcal{R}_{\beta,M}(\psi)$, which is discussed later.

With these preliminaries at hand, we propose to find ψ (as well as ϕ_1, ϕ_2) as the solution to the following minimization problem:

$$\begin{aligned}\min_{\substack{0\le\phi_1\le 1,\\ 0\le\phi_2\le 1,\\ \psi}} \Bigg\{ \mathcal{F}(\phi_1,\phi_2,\psi) = \int_\Omega |\nabla\phi_1| + \int_\Omega |\nabla\phi_2| + \mathcal{R}_{\beta,M}(\psi) \\ + \frac{\lambda}{2}|\alpha|^2 \int_\Omega \Bigg[\phi_2(x)\phi_1(x)\left(h(x) - \frac{\psi(x)}{\alpha_1}\right)^2 \\ + \phi_2(x)(1-\phi_1(x))\left(h(x) - \frac{\psi(x)}{\alpha_3}\right)^2 \\ + (1-\phi_2(x))\left(h(x) - \frac{\psi(x)}{\alpha_2}\right)^2 \Bigg] dx \Bigg\}\end{aligned} \tag{4.1}$$

with $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. In fact, the parameter $\frac{\lambda}{2}|\alpha|^2$ can be replaced simply by λ; however, we kept the factor $|\alpha|^2$ because (4.1) was derived initially from a vectorial setting where this factor appeared. Keeping the factor has no effect on the answer.

The minimization problem (4.1) can be understood as follows. For a fixed ψ, the objective function in (4.1) attempts to find a partition represented by ϕ_1, ϕ_2 such that ψ/α_i approximates h closely in each subset of the partition. The first two terms in the functional regularize ϕ_1 and ϕ_2. Further, for fixed ϕ_1 and ϕ_2, the functional in (4.1) attempts to find a regularized ψ that further minimizes the difference $(h - \psi/\alpha_i)$ in in each subset of the partition.

When it comes to $\mathcal{R}_{\beta,M}$, a regularizer for ψ, to ensure that ψ is a slowly varying smooth function, we require ψ to be a bounded H^1 function by using

$$\mathcal{R}_{\beta,M}(\psi) = \beta \int_\Omega |\nabla\psi|^2 + \chi_M(\psi),$$

where $\beta > 0$ and

$$\chi_M(\psi) = \begin{cases} 0 & \text{if } \|\psi\|_\infty \le M, \\ \infty & \text{otherwise} \end{cases}$$

for some $0 < M < \infty$ since we may well expect that a slowly varying smooth function over a bounded region should be bounded. Note that $\beta > 0$ controls the variation of the shape of the function ψ. An important observation when choosing $\mathcal{R}_{\beta,M}$ is that the solution function ψ should be independent of scaling h by a positive constant. In other words, our choice of $\mathcal{R}_{\beta,M}$ with $M = \infty$ above confirms that with $\phi_1, \ldots, \phi_{K-1}$ fixed, ψ^* minimizes the functional in (4.1) with input h if and only if $\gamma\psi^*$ minimizes the same functional with input γh, $\gamma > 0$. Our choice of $\mathcal{R}_{\beta,M}$ is appropriate in this sense. Since the boundedness of h implies the boundedness of ψ^* as will be seen later, $\mathcal{R}_{\beta,M}$ with $M = \infty$ is of interest in most applications.

To understand the minimization problem (4.1) further, we introduce three supplemental functionals:

$$\mathcal{F}^1_{\phi_2,\psi}(u) = \int_\Omega |\nabla u| + \frac{\lambda}{2}|\alpha|^2 \int_\Omega \phi_2 \left[\left(h - \frac{\psi}{\alpha_1} \right)^2 - \left(h - \frac{\psi}{\alpha_3} \right)^2 \right] u,$$

$$\mathcal{F}^2_{\phi_1,\psi}(u) = \int_\Omega |\nabla u| + \frac{\lambda}{2}|\alpha|^2 \int_\Omega \left\{ \left[\phi_1 \left(h - \frac{\psi}{\alpha_1} \right)^2 + (1-\phi_1) \left(h - \frac{\psi}{\alpha_3} \right)^2 \right] - \left(h - \frac{\psi}{\alpha_2} \right)^2 \right\} u,$$

$$\mathcal{F}^3_{\phi_1,\phi_2}(u) = \mathcal{R}_{\beta,M}(u) + \frac{\lambda}{2}|\alpha|^2 \int_\Omega \left[\left(\frac{1}{\alpha_1^2}\phi_2\phi_1 + \frac{1}{\alpha_3^2}\phi_2(1-\phi_1) + \frac{1}{\alpha_2^2}(1-\phi_2) \right) u^2 \right.$$
$$\left. - 2h \left(\frac{1}{\alpha_1}\phi_2\phi_1 + \frac{1}{\alpha_3}\phi_2(1-\phi_1) + \frac{1}{\alpha_2}(1-\phi_2) \right) u + h^2 \right].$$

Then, we can observe the following:

$$\operatorname*{argmin}_{0\le\phi_1\le 1} \mathcal{F}(\phi_1,\phi_2,\psi) = \operatorname{argmin}_{0\le\phi_1\le 1} \mathcal{F}^1_{\phi_2,\psi}(\phi_1), \tag{4.2}$$

$$\operatorname*{argmin}_{0\le\phi_2\le 1} \mathcal{F}(\phi_1,\phi_2,\psi) = \operatorname{argmin}_{0\le\phi_2\le 1} \mathcal{F}^2_{\phi_1,\psi}(\phi_2), \tag{4.3}$$

and

$$\operatorname*{argmin}_{\psi} \mathcal{F}(\phi_1,\phi_2,\psi) = \operatorname{argmin}_{\psi} \mathcal{F}^3_{\phi_1,\phi_2}(\psi). \tag{4.4}$$

Theorem 4.2. *Let $h \in L^2(\Omega)$. Then, there exist binary functions ϕ_1^*, ϕ_2^* taking values 0 and 1 and an H^1 function ψ^* such that $(\phi_1^*, \phi_2^*, \psi^*)$ solves the minimization problem (4.1), in other words,*

$$\mathcal{F}(\phi_1^*, \phi_2^*, \psi^*) \le \mathcal{F}(\phi_1, \phi_2, \psi)$$

for any $0 \le \phi_1 \le 1$, $0 \le \phi_2 \le 1$, and $\psi \in H^1(\Omega)$.

Proof. The proof uses a standard technique in optimization theory. Since the functional $\mathcal{F}$ is bounded from below, we may choose a minimizing sequence

$$\{(\phi_1^n, \phi_2^n, \psi^n)\}_{n=1}^\infty.$$

Note that $\{\phi_1^n\}_{n=1}^\infty$ and $\{\phi_2^n\}_{n=1}^\infty$ are bounded sequences in BV and $\{\psi^n\}_{n=1}^\infty$ is a bounded sequence in $H^1(\Omega)$ with $\sup_n \|\psi^n\|_\infty \leq M$. This makes it possible for us to extract subsequences that possess weak-$*$ or weak limits, that is, there is a subsequence $\{(\phi_1^{n_k}, \phi_2^{n_k}, \psi^{n_k})\}_{k=1}^\infty$ such that as $k \to \infty$,

$$\phi_1^{n_k} \to u_1^* \quad \text{pointwise a.e.}, \quad \phi_2^{n_k} \to u_2^* \quad \text{pointwise a.e.}, \quad \psi^{n_k} \to \psi^* \quad \text{pointwise a.e.}$$

and

$$\int_\Omega |\nabla u_1^*| \leq \liminf_{k\to\infty} \int_\Omega |\nabla \phi_1^{n_k}| \quad \text{and} \quad \int_\Omega |\nabla u_2^*| \leq \liminf_{k\to\infty} \int_\Omega |\nabla \phi_2^{n_k}|$$

and

$$\psi^{n_k} \rightharpoonup \psi^* \quad \text{in } H^1(\Omega) \text{ as } k \to \infty.$$

Then, we can easily obtain that $\|\psi^*\|_\infty \leq M$ and

$$\mathcal{F}(u_1^*, u_2^*, \psi^*) \leq \lim_{k\to\infty} \mathcal{F}(\phi_1^{n_k}, \phi_2^{n_k}, \psi^{n_k}),$$

which means that (u_1^*, u_2^*, ψ^*) solves (4.1). From (4.2) and [5], [13], we can see that

$$\mathcal{F}(u_1^*, u_2^*, \psi^*) = \mathcal{F}(\phi_1^*, u_2^*, \psi^*),$$

where ϕ_1^* is a binary function taking values 0 and 1 such that $\mathcal{F}^1_{u_2^*,\psi^*}(\phi_1^*) = \min_{0\leq\phi\leq 1} \mathcal{F}^1_{u_2^*,\psi^*}(\phi)$. With the same argument, (4.3) and [5], [13] imply

$$\mathcal{F}(\phi_1^*, u_2^*, \psi^*) = \mathcal{F}(\phi_1^*, \phi_2^*, \psi^*),$$

where ϕ_2^* is a binary function taking values 0 and 1 such that $\mathcal{F}^2_{\phi_1^*,\psi^*}(\phi_2^*) = \min_{0\leq\phi\leq 1} \mathcal{F}^2_{\phi_1^*,\psi^*}(\phi)$. Therefore, $(\phi_1^*, \phi_2^*, \psi^*)$ solves (4.1) and ϕ_1^*, ϕ_2^* are binary functions taking values 0 and 1 and ψ^* is an $H^1(\Omega)$ function. ■

Variational Algorithm 1. *The minimization problem in* (4.1) *can be solved via the following procedure.*

1. *Set* $n = 0$, *and initialize* ϕ_1^0 *and* ϕ_2^0 *as characteristic functions.*
2. *Find* (u^*, ω^*) *such that* $0 \leq u^* \leq 1$,

$$\mathcal{F}(u^*, \phi_2^n, \omega^*) = \min_{0\leq u\leq 1, \psi} \mathcal{F}(u, \phi_2^n, \omega), \tag{4.5}$$

   and set $\phi_1^{n+1} = u^*$.
3. *Find* (v^*, ψ^*) *such that* $0 \leq v^* \leq 1$,

$$\mathcal{F}(\phi_1^{n+1}, v^*, \psi^*) = \min_{0\leq v\leq 1, \psi} \mathcal{F}(\phi_1^{n+1}, v, \psi), \tag{4.6}$$

   and set $\phi_2^{n+1} = v^*$ *and* $\psi^{n+1} = \psi^*$.
4. $n \leftarrow n + 1$ *and go to* 2.

We would like to point out that ω^* obtained in (4.5) is not explicitly used for the next step in (4.6). Nevertheless, it is important to compute ω^* for a proper u^* in (4.5). In the numerical computations, ω^* is taken as an initial guess for computing ψ^* in (4.6). When solving (4.5), we will alternate between

$$\min_{0\le u\le 1} \mathcal{F}(u,\phi_2^n,\omega) \quad \Leftrightarrow \quad \min_{0\le u\le 1} \mathcal{F}^1_{\phi_2^n,\omega}(u)$$

and

$$\min_{\omega} \mathcal{F}(u,\phi_2^n,\omega) \quad \Leftrightarrow \quad \min_{\omega} \mathcal{F}^3_{u,\phi_2^n}(\omega).$$

Similarly, when solving (4.6), we will also alternate between

$$\min_{0\le v\le 1} \mathcal{F}(\phi_1^{n+1},v,\psi) \quad \Leftrightarrow \quad \min_{0\le v\le 1} \mathcal{F}^2_{\phi_1^{n+1},\psi}(v)$$

and

$$\min_{\psi} \mathcal{F}(\phi_1^{n+1},v,\psi) \quad \Leftrightarrow \quad \min_{\psi} \mathcal{F}^3_{\phi_1^{n+1},v}(\psi).$$

Therefore, we will analyze the three subproblems corresponding to the three supplemental functionals

$$\min_{0\le u\le 1} \mathcal{F}^1_{\varphi_1,\varphi_2}(u) \quad \text{and} \quad \min_{0\le u\le 1} \mathcal{F}^2_{\varphi_1,\varphi_2}(u) \quad \text{and} \quad \min_{u} \mathcal{F}^3_{\varphi_1,\varphi_2}(u).$$

From now on, the sequence $\{(\phi_1^n,\phi_2^n,\psi^n)\}_{n=1}^{\infty}$ will be referred to as the one obtained by the *Variational Algorithm*. The following theorems from [5] and [13] can help us minimize $\mathcal{F}^1_{\phi_2^n,\psi}$ and $\mathcal{F}^2_{\phi_1^{n+1},\psi}$.

Theorem 4.3. *Let $h\in L^2(\Omega)$. Let ϕ^* be a minimizer of*

$$\min_{0\le\phi\le 1}\left\{\int_\Omega |\nabla\phi| + \lambda\int_\Omega h(x)\phi(x)dx\right\}. \tag{4.7}$$

Then for a.e. $\mu\in[0,1]$, the function 1_{Σ_μ} is also a minimizer of (4.7), where

$$\Sigma_\mu = \{x\in\Omega : \phi^*(x) > \mu\}.$$

Proof. This is the same as Theorem 2 of [5]. ∎

It is easy to see that $\phi^*=0$ is a unique minimizer of (4.7) if $h\ge 0$ a.e. and $h\ne 0$. Hence, we are only interested in a datum h such that

$$\{x\in\Omega : h(x) < 0\}$$

has positive Lebesgue measure.

Theorem 4.4 (Theorem 4 of [13]). *Let $h\in L^2(\Omega)$ be such that $\{x\in\Omega : h(x)<0\}$ has positive Lebesgue measure. We consider the following problem:*

$$\min_{\omega}\left\{\int_\Omega |\nabla\omega| + \int_\Omega \left(\omega(x)+\frac{\lambda}{2}h(x)\right)^2 dx\right\}. \tag{4.8}$$

Then, there exists a unique minimizer ω^* *of* (4.8). *Moreover,* $\Sigma_* = \{\omega^* > 0\}$ *is a minimizer of*

$$\min_{\Sigma\subset\Omega}\left\{\int_\Omega |\nabla 1_\Sigma| + \lambda\int_\Sigma h(x)dx\right\}. \tag{4.9}$$

In addition, (4.9) *has a unique minimizer if and only if* $\{\omega^* = 0\}$ *has Lebesgue measure* 0. *Otherwise,* $\{\omega^* > 0\}$ *and* $\{\omega^* \geq 0\}$ *are the minimal and the maximal solutions of* (4.9).

Theorems 4.3 and 4.4 solve (4.9) and imply that the convex constrained problem (4.7) can be solved by the strictly convex unconstrained problem (4.8). We can see from the examples provided in [13] that (4.7) presents uncertainty in computing a minimizer in the sense that we cannot tell which minimizer would be computed when there is more than one minimizer. On the other hand, (4.8) always presents the minimal and the maximal solutions of (4.9). Therefore, any minimization problem of the form (4.7) will be approached via (4.8) throughout the paper, and we will compute the minimal solution of (4.9) as far as (4.8) is concerned. This is summarized in Proposition 4.5 below.

Proposition 4.5. *Let* $h \in L^2(\Omega)$ *and* $\psi \in H^1(\Omega)$ *with* $\|\psi\|_\infty \leq M$ *and* $\phi_1^{n+1}, \phi_2^n \in BV(\Omega)$ *with* $0 \leq \phi_1^{n+1}, \phi_2^n \leq 1$. *Then,*

$$\min_{0\leq u\leq 1} \mathcal{F}^1_{\phi_2^n,\psi}(u) \tag{4.10}$$

can be solved by (4.8) *with* λ *and* h *replaced by* $\frac{\lambda}{2}|\alpha|^2$ *and*

$$\phi_2^n\left[\left(h-\frac{\psi}{\alpha_1}\right)^2 - \left(h-\frac{\psi}{\alpha_3}\right)^2\right],$$

respectively. Moreover, the minimizer u^* *of* (4.10) *obtained via Theorem* 4.4 *is a characteristic function* $1_{\Sigma^*} \in BV(\Omega)$ *that is also a solution of*

$$\min_{\Sigma\subset\Omega}\left\{\mathrm{P}(\Sigma;\Omega) + \frac{\lambda}{2}|\alpha|^2\int_\Sigma \phi_2^n\left(h-\frac{\psi}{\alpha_1}\right)^2 + \frac{\lambda}{2}|\alpha|^2\int_{\Omega\setminus\Sigma}\phi_2^n\left(h-\frac{\psi}{\alpha_3}\right)^2\right\}.$$

Likewise,

$$\min_{0\leq\phi_2\leq 1} \mathcal{F}^2_{\phi_1^{n+1},\psi}(\phi_2) \tag{4.11}$$

can be solved in the same way using (4.8) *with* λ *and* h *replaced by* $\frac{\lambda}{2}|\alpha|^2$ *and*

$$\left[\phi_1^{n+1}\left(h-\frac{\psi}{\alpha_1}\right)^2 + (1-\phi_1^{n+1})\left(h-\frac{\psi}{\alpha_3}\right)^2\right] - \left(h-\frac{\psi}{\alpha_2}\right)^2,$$

respectively. Then, the minimizer u^* *of* (4.11) *obtained via Theorem* 4.4 *is also a characteristic function* $1_{\Sigma^*} \in BV(\Omega)$ *that is a solution of*

$$\min_{\Sigma\subset\Omega}\left\{\mathrm{P}(\Sigma;\Omega) + \frac{\lambda}{2}|\alpha|^2\int_\Sigma\left[\phi_1^{n+1}\left(h-\frac{\psi}{\alpha_1}\right)^2 + (1-\phi_1^{n+1})\left(h-\frac{\psi}{\alpha_3}\right)^2\right] + \frac{\lambda}{2}|\alpha|^2\int_{\Omega\setminus\Sigma}\left(h-\frac{\psi}{\alpha_2}\right)^2\right\}.$$

Proof. All we need to check is that

$$\phi_2^n\left[\left(h-\frac{\psi}{\alpha_1}\right)^2-\left(h-\frac{\psi}{\alpha_3}\right)^2\right]\in L^2(\Omega)$$

and

$$\left[\phi_1^{n+1}\left(h-\frac{\psi}{\alpha_1}\right)^2+(1-\phi_1^{n+1})\left(h-\frac{\psi}{\alpha_3}\right)^2\right]-\left(h-\frac{\psi}{\alpha_2}\right)^2\in L^2(\Omega),$$

which are equivalent, respectively, to

$$2\left(\frac{1}{\alpha_3}-\frac{1}{\alpha_1}\right)h\psi+\left(\frac{1}{\alpha_1^2}-\frac{1}{\alpha_3^2}\right)\psi^2\in L^2(\Omega)$$

and

$$2\left[\frac{1}{\alpha_2}-\left(\frac{\phi_1^{n+1}}{\alpha_1}+\frac{(1-\phi_1^{n+1})}{\alpha_3}\right)\right]h\psi+\left(\frac{\phi_1^{n+1}}{\alpha_1^2}+\frac{(1-\phi_1^{n+1})}{\alpha_3^2}-\frac{1}{\alpha_2^2}\right)\psi^2\in L^2(\Omega).$$

These can be easily confirmed by $\psi\in L^\infty(\Omega)$ and $h\in L^2(\Omega)$ and $0\le\phi_1^{n+1},\phi_2^n\le 1$. The rest can be easily observed from Theorems 4.3 and 4.4. ∎

We note that for each $n=1,2,\dots,$ ϕ_1^n and ϕ_2^n are obtained by minimizing $\mathcal{F}^1$ and $\mathcal{F}^2$ using the method described in Theorem 4.4, which implies that they are characteristic functions that are minimal solutions of the corresponding problems in the form of (4.9) in our algorithm. So we may assume that ϕ_1^n and ϕ_2^n are characteristic functions for all $n\ge 1$. Then, ψ^{n+1} satisfies

$$\begin{aligned}\mathcal{F}(\phi_1^{n+1},\phi_2^{n+1},\psi^{n+1})&=\min_{\psi}\mathcal{F}^3_{\phi_1^{n+1},\phi_2^{n+1}}(\psi)\\&=\min_{\psi}\left\{\mathcal{R}_{\beta,M}(\psi)+\frac{\lambda}{2}|\alpha|^2\int_\Omega(\Phi^{n+1}(x)\psi(x)-h(x))^2dx\right\},\end{aligned}$$

where

$$\begin{aligned}\Phi^{n+1}&=\frac{1}{\alpha_1}\phi_2^{n+1}\phi_1^{n+1}+\frac{1}{\alpha_3}\phi_2^{n+1}(1-\phi_1^{n+1})+\frac{1}{\alpha_2}(1-\phi_2^{n+1})\\&=\begin{cases}\frac{1}{\alpha_1} & \text{on } \{\phi_1^{n+1}=1,\ \phi_2^{n+1}=1\},\\ \frac{1}{\alpha_3} & \text{on } \{\phi_1^{n+1}=0,\ \phi_2^{n+1}=1\},\\ \frac{1}{\alpha_2} & \text{on } \{\phi_2^{n+1}=0\}.\end{cases}\end{aligned}$$

The following proposition applies to most applications and it shows that the condition

$$\|\psi^{n+1}\|_\infty\le M$$

does not impose any computational burden on the algorithm.

Proposition 4.6. *Let $h \in L^\infty(\Omega)$. If $M > \alpha_3\|h\|_\infty$, then the solution ψ^{n+1} of the problem*

$$\min_{\psi} \mathcal{F}^3_{\phi_1^{n+1},\phi_2^{n+1}}(\psi) = \min_{\psi}\left\{\mathcal{R}_{\beta,M}(\psi) + \frac{\lambda}{2}|\alpha|^2\int_\Omega (\Phi^{n+1}(x)\psi(x) - h(x))^2 dx\right\}$$

satisfies

$$\|\psi^{n+1}\|_\infty \le \alpha_3\|h\|_\infty.$$

Proof. Simply speaking, the maximum principle applies, i.e.,

$$\|\Phi^{n+1}\psi^{n+1}\|_\infty \le \|h\|_\infty.$$

Since $\alpha_1 < \alpha_2 < \alpha_3$, we obtain that

$$\|\psi^{n+1}\|_\infty \le \alpha_3\|h\|_\infty. \qquad \blacksquare$$

Proposition 4.6 implies that $\mathcal{R}_{\beta,M}(\psi)$ and

$$\beta\int_\Omega |\nabla\psi(x)|^2 dx$$

play the same role with $M > \alpha_3\|h\|_\infty$. The following convergence result also holds.

Theorem 4.7. *The sequence $\{(\phi_1^n, \phi_2^n, \psi^n)\}_{n=1}^\infty$ obtained from the Variational Algorithm satisfies*

$$\mathcal{F}(\phi_1^{n+1}, \phi_2^{n+1}, \psi^{n+1}) \le \mathcal{F}(\phi_1^n, \phi_2^n, \psi^n) \quad \textit{for all } n = 1, 2, \ldots,$$

and possesses a limit point $(\phi_1^*, \phi_2^*, \psi^*)$ such that

$$\mathcal{F}(\phi_1^*, \phi_2^*, \psi^*) \le \inf_n \mathcal{F}(\phi_1^n, \phi_2^n, \psi^n).$$

Proof. Note that for each $n = 1, 2, \ldots$ we have from the Variational Algorithm that

$$\mathcal{F}(\phi_1^{n+1}, \phi_2^{n+1}, \psi^{n+1}) \le \mathcal{F}(\phi_1^{n+1}, \phi_2^n, \omega^*) \le \mathcal{F}(\phi_1^n, \phi_2^n, \psi^n). \qquad \blacksquare$$

Remark 1. Just as with other alternating minimization methods, it is difficult to see whether a limit point from Theorem 4.7 can serve as a minimizer in Theorem 4.2. However, we can make sure that our algorithm produces a convergent sequence decreasing the functional $\mathcal{F}$.

5. Theory for the general case ($K \ge 2$). Having dealt with the $K = 3$ case, we now briefly mention the general $K \ge 2$ case. The same analysis as the $K = 3$ case can be carried over to any $K \ge 2$.

Theorem 5.1. *Let $\Omega \subset \mathbb{R}^N$ be open and connected, $N \ge 1$. Suppose that a given function h on Ω is a product of two unknown functions f and g through $h = f \cdot g$. Suppose g is positive and continuous in Ω and f is positive and piecewise constant with $K \ge 2$, i.e.,*

$$f = \mu_1 1_{\Sigma_1} + \cdots + \mu_K 1_{\Sigma_K},$$

where $\mu_1 > \cdots > \mu_K$ *and the partition* $\Omega = \Sigma_1 \cup \cdots \cup \Sigma_K$ *is unknown. Suppose that the following* $K-1$ *ratios*

$$\rho_i = \frac{\mu_i}{\mu_{i+1}},\ i = 1, 2, \ldots, K-1,$$

are given. Then, there exist $\alpha_1, \ldots, \alpha_K$ *in* $\mathbb{R}$ *depending only on* $\rho_1, \ldots, \rho_{K-1}$ *such that the set* X *defined by*

$$\Big\{u : \Omega \to \mathbb{R} : u(x) \in \{\alpha_1 h(x), \ldots, \alpha_K h(x)\}\Big\},$$

contains a continuous function $\tilde{g}$*, which is a constant multiple of the unknown function* g*. Moreover,* $\tilde{g}$ *is a unique continuous function in* X *and* $h/\tilde{g}$ *is a constant multiple of the unknown function* f*.*

The constants α_i in the above theorem are determined for $\rho_1, \ldots, \rho_{K-1}$ by

$$\alpha_i = \left(\prod_{j\neq i} \frac{\mu_j}{\mu_i}\right)^{\frac{1}{K}} = \left(\left(\prod_{j<i}\prod_{k=j}^{i-1} \rho_k\right)\left(\prod_{j>i}\prod_{k=i}^{j-1} \rho_k^{-1}\right)\right)^{\frac{1}{K}}$$

and the proof is the same as that of Theorem 4.1 In addition, the minimization problem (4.1) can be extended to any $K \geq 2$ as follows:

$$\min_{\substack{0\leq\phi_1,\ldots,\phi_{K-1}\leq 1,\\ \psi}} \mathcal{F}(\phi_1, \ldots, \phi_{K-1}, \psi),$$

where if $K \geq 2$ is an even number, then

$$\begin{aligned}
\mathcal{F}(\phi_1, \ldots, \phi_{K-1}, \psi) &= \sum_{i=1}^{K-1} \int_\Omega |\nabla \phi_i| + \mathcal{R}_{\beta,M}(\psi) \\
&+ \frac{\lambda}{2}|\alpha|^2 \left\{ \int_\Omega \left[\left\{ \sum_{i=1}^{K/2-1} \left(\prod_{j=K/2+1}^{K/2+i-1} (1-\phi_j(x)) \right) \phi_{K/2+i}(x) \right.\right.\right. \\
&\qquad \left.\left.\left. \times \left(\phi_i(x) \left(h(x) - \frac{\psi(x)}{\alpha_i} \right)^2 + (1-\phi_i(x)) \left(h(x) - \frac{\psi(x)}{\alpha_{K+1-i}} \right)^2 \right) \right\} \right. \\
&\qquad + \left(\prod_{j=K/2+1}^{K-1} (1-\phi_j(x)) \right) \left(\phi_{K/2}(x) \left(h(x) - \frac{\psi(x)}{\alpha_{K/2}} \right)^2 \right. \\
&\qquad \left.\left. \left. + (1-\phi_{K/2}(x)) \left(h(x) - \frac{\psi(x)}{\alpha_{K/2+1}} \right)^2 \right) \right] dx \right\},
\end{aligned}$$

whereas if $K > 2$ is an odd number, then

$$\mathcal{F}(\phi_1, \ldots, \phi_{K-1}, \psi) = \sum_{i=1}^{K-1} \int_\Omega |\nabla \phi_i| + \mathcal{R}_{\beta,M}(\psi)$$
$$+ \frac{\lambda}{2}|\alpha|^2 \left\{ \int_\Omega \left[\left\{ \sum_{i=1}^{(K-1)/2} \left(\prod_{j=(K+1)/2}^{(K-1)/2+i-1} (1 - \phi_j(x)) \right) \phi_{(K+1)/2+i-1}(x) \right.\right.$$
$$\left. \times \left(\phi_i(x) \left(h(x) - \frac{\psi(x)}{\alpha_i} \right)^2 + (1 - \phi_i(x)) \left(h(x) - \frac{\psi(x)}{\alpha_{K+1-i}} \right)^2 \right) \right\}$$
$$\left. + \left(\prod_{j=(K+1)/2}^{K-1} (1 - \phi_j(x)) \right) \left(h(x) - \frac{\psi(x)}{\alpha_{(K+1)/2}} \right)^2 \right] dx \right\},$$

where we use the convention that $\prod_{j=m}^{k} s_j = 1$ if $k < m$.

For the purpose of illustration, we consider the case when $K > 1$ is an odd number. This applies to the case of an even number $K > 1$ in exactly the same way. For

$$f = \mu_1 1_{\Sigma_1} + \cdots + \mu_K 1_{\Sigma_K},$$

where $\mu_1 > \cdots > \mu_K$ and the partition is $\Omega = \Sigma_1 \cup \cdots \cup \Sigma_K$, we would like to pair up Σ_i and Σ_{K+1-i}, $i = 1, 2, \ldots, (K-1)/2$, and represent them by

$$\Sigma_i = \left\{ x \in \Omega : \prod_{j=(K+1)/2}^{(K-1)/2+i-1} (1 - \phi_j(x)) = 1 \text{ and } \phi_{(K+1)/2+i-1}(x) = 1 \right\} \cap \{x \in \Omega : \phi_i(x) = 1\},$$

and

$$\Sigma_{K+1-i} = \left\{ x \in \Omega : \prod_{j=(K+1)/2}^{(K-1)/2+i-1} (1 - \phi_j(x)) = 1 \text{ and } \phi_{(K+1)/2+i-1}(x) = 1 \right\}$$
$$\cap \{x \in \Omega : \phi_i(x) = 0\}.$$

As for $i = (K+1)/2$, we represent $\Sigma_{(K+1)/2}$ by

$$\Sigma_{(K+1)/2} = \left\{ x \in \Omega : \prod_{j=(K+1)/2}^{K-1} (1 - \phi_j(x)) = 1 \right\}.$$

6. Numerical techniques. We now provide specific numerical algorithms to solve (4.2), (4.3), (4.4). Among the many available algorithms, we use the algorithm by Chambolle [3] for (4.2) and (4.3) and the Gauss–Seidel method provided in [11] for (4.4). Details are given below. We chose Chambolle's algorithm [3] because it was reported in [13] that this algorithm is simple and efficient for (4.9) due to the particular stopping criterion that we

impose. Any other efficient algorithms for solving (4.8) should work. One can use those algorithms provided in [4] for all of (4.2), (4.3), (4.4). However, in our experiments, we observed better performances in using [3] than [4] with comparable computational speed as was reported in [13]. Note that (4.5) is approached via alternation between (4.2) and (4.4) and (4.6) is approached via alternation between (4.3) and (4.4).

In fact, we can borrow the detailed setting for Chambolle's algorithm with the new stopping criterion from [13] to solve (4.9): For a given $\epsilon > 0$ and for each $k = 1, 2, \ldots$, we compute

$$p_{i,j,l}^{k+1} = \frac{p_{i,j,l}^{k} + \tau(\nabla(\operatorname{div}(p^k) + \lambda h))_{i,j,l}}{1 + +\tau|(\nabla(\operatorname{div}(p^k) + \lambda h))_{i,j,l}|} \tag{6.1}$$

and stop when

$$|\Sigma_{k+1}\Delta\Sigma_k| = |\Sigma_{k+1} \setminus \Sigma_k| + |\Sigma_k \setminus \Sigma_{k+1}| < \epsilon,$$

where

$$\Sigma_k = \left\{ (i,j,l) : \omega_{i,j,l}^k = -\frac{\lambda}{2} h_{i,j,l} - \frac{1}{2}(\operatorname{div}(p^k))_{i,j,l} > 0 \right\}.$$

For all the experiments, we used $\epsilon = 1$, i.e., we stopped the iteration when $\Sigma_{k+1} \equiv \Sigma_k$, which was considered to be exact recovery in [13], to make sure of the accuracy of our algorithm. In fact, we checked this stopping criterion at every 10th iteration. And we simply replace λ and h in (6.1) by those given in Proposition 4.5 accordingly to compute (4.2) and (4.3).

As for (4.4), since the input data h are bounded in practice, Proposition 4.6 allows us to solve

$$\min_{\psi} \left\{ \beta \int_\Omega |\nabla\psi|^2 + \frac{\tilde{\lambda}}{2} \int_\Omega (\Phi(x)\psi(x) - h(x))^2 dx \right\}, \tag{6.2}$$

where

$$\begin{aligned} \Phi &= \frac{1}{\alpha_1}\phi_2\phi_1 + \frac{1}{\alpha_3}\phi_2(1-\phi_1) + \frac{1}{\alpha_2}(1-\phi_2) \\ &= \begin{cases} \frac{1}{\alpha_1} & \text{on } \{\phi_1 = 1,\ \phi_2 = 1\}, \\ \frac{1}{\alpha_3} & \text{on } \{\phi_1 = 0,\ \phi_2 = 1\}, \\ \frac{1}{\alpha_2} & \text{on } \{\phi_2 = 0\} \end{cases} \end{aligned}$$

with $\tilde{\lambda} = \lambda|\alpha|^2$. Note that ϕ_1, ϕ_2 that we use for (4.4) will be obtained from (4.2) and (4.3), i.e., they are binary functions, which permits Φ to take three values as above. Strict convexity of (6.2) implies that there exists a unique minimizer ψ^* of (6.2) that satisfies

$$(\tilde{\lambda}\Phi^2 - 2\beta\Delta)\psi^* = \tilde{\lambda} h \cdot \Phi. \tag{6.3}$$

The three-dimensional version of the Gauss–Seidel method in [11] for (6.3) reads as follows:

with $\omega = \Phi^2$ and $f = h \cdot \Phi$,

$$\psi^{k+1} = \{G^k_{i,j,l}\} = \left\{ \frac{2\beta}{\tilde{\lambda}\omega_{i,j} + 12\beta} \left(\psi^k_{i+1,j,l} + \psi^k_{i-1,j,l} + \psi^k_{i,j+1,l} + \psi^k_{i,j-1,l} + \psi^k_{i,j,l+1} + \psi^k_{i,j,l-1} \right) + \frac{\tilde{\lambda}}{\tilde{\lambda}\omega_{i,j} + 12\beta} f \right\}.$$

We will use three-dimensional masks for the brain data sets in all the experiments to describe the boundaries and consider Dirichlet boundary conditions. We will stop this iteration when

$$\|\psi^{k+1} - \psi^k\|_2^2 = \sum_{i,j,l} \left| G^k_{i,j,l} - G^{k-1}_{i,j,l} \right|^2 < \epsilon_1.$$

The algorithm in [4] can also serve as an alternative to solve (4.4). The parameters for our algorithm are set to $\lambda = 0.02, \tau = 0.1, \beta = 4, \epsilon_1 = 0.5$. The most influential parameter turns out to be β, in fact, $\frac{\lambda}{\beta}$, which determines the smoothness of the smooth and slowly varying function. When it comes to real MR images, we found that $\beta = 4$ provided reasonable and reliable results when all the other parameters were fixed. In real data sets, $\beta \approx 2$ started to capture local structures whereas $\beta \approx 6$ resulted in oversmoothing. For synthetic data, $\beta \approx 1$ worked better than $\beta = 4$. We believe that this is due to the fact that the synthetic data sets are prepared from a truly piecewise constant function f. Moreover, the smaller β is, the faster our algorithm runs. As for alternating between (4.5) and (4.6), we iterate (4.2) and (4.4) three times for (4.5) and iterate (4.3) and (4.4) three times for (4.6), which worked well enough in practice. During these iterations, we always check to make sure that the functional is decreasing at every iteration. As for terminating the Variational Algorithm, we run it until it reaches a steady state solution when dealing with synthetic MR images, where our theory is truly applicable. With real MR images, we can terminate the algorithm at $n = 2$; in other words, the smooth and slowly varying function that we obtained in our experiments with real MR images was all ψ^2 since we did not observe any significant improvement in the subsequent iterations. It is also important to have a good initial guess for ϕ^0_1 and ϕ^0_2. We always used ϕ^0_1 densely defined in Ω with $\phi^0_2 \equiv 1$. More detailed explanations will come in section 8.

7. Adaptation. What has been discussed so far is based on the assumption that accurate estimates of the the ratios (ρ_1, ρ_2) are available. As we show below in section 8, the algorithm is not sensitive to this assumption, especially with real MR images. Nevertheless, to further improve robustness of the algorithm, we propose an adaptive version where the ratios (ρ_1, ρ_2) are updated in an outer loop.

Adaptive Variational Algorithm 1. *Given estimated ratios (ρ_1, ρ_2), do the following.*

1. *Use the Variational Algorithm with estimated (ρ_1, ρ_2) to compute ψ. As a by-product, obtain estimates of the indicator functions $1_{\Sigma_1}, 1_{\Sigma_2}$, and 1_{Σ_3} of the three regions.*
2. *We divide the input h by ψ to estimate the underlying true signal f. Then obtain the three average values $\mu_1 > \mu_2 > \mu_3$ in the estimated regions.*
3. *Set $\rho_1 = \mu_1/\mu_2,\ \rho_2 = \mu_2/\mu_3$.*

This automated procedure will be adopted for nonuniformity correction with real MR images in the next section, where accurate estimates of the ratios are not available.

8. Simulations and experiments with real MR images.

8.1. Overview. We tested the algorithm with simulations and real MR images, and in this section, we report the results of these tests. In the simulations, the data were generated according to the model (1.1), so that the "ground truth" inhomogeneity was known. The accuracy of the algorithm was assessed by the following criteria.

1. Ideally, the estimated nonuniformity ψ is proportional to g, the true nonuniformity. In practice, the two may depart from strict proportionality because of noise, numerical error, or the algorithm being trapped in a local minimum. We measured departure from proportionality in the following way: First, we normalized (scaled) g/ψ to have a maximum value of 1. Then, we created a histogram of the normalized g/ψ values in the brain region. Ideally, this histogram should have a spike at the value of 1, but loss of proportionality will cause the spike to spread out. We use the variance of the histogram of normalized g/ψ as a performance measure. Smaller variances imply better performance. We refer to this variance as the *normalized variance*.
2. We use some secondary performance measures as well. These are not as important as the normalized variance, but they do assess whether the assumptions of the theory hold. Because the algorithm depends on segmenting the image, we measured the accuracy of the segmentation relative to the ground truth in the simulations. As another secondary performance measure, we compared the ρ values in the final segmentation and compared them to the ground truth ρ values.

The simulation experiments were carried out in three steps. First, the algorithm was tested on a noise-free data set to test the intrinsic accuracy of the algorithm. Next the algorithm was tested on noisy data. In both of these simulations the ρ values were assumed to be known. In the final simulation, we discarded this assumption, and used the adaptive algorithm which was initialized at incorrect vales of ρ.

8.2. Simulations. The synthetic data set we used was the simulated scan with gradient available from the Center for Morphometric Analysis at Massachusetts General Hospital at http://www.cma.mgh.harvard.edu/ibsr/. The original size of the data is $256 \times 256 \times 55$ (voxel size), which we reduced to $118 \times 116 \times 47$ by trimming out the background region and by discarding the first and the last few slices. Then, we added one extra slice at each of the two ends of the brain region in each dimension to take care of the boundary conditions and prepared a mask for the brain region as well. The GM/WM/CSF regions were very clearly separable using their intensity values, and this produced the segmentation ground truth for the regions $\Sigma_1, \Sigma_2, \Sigma_3$. The three intensity values in $\Sigma_1, \Sigma_2, \Sigma_3$ were $\mu_1 = 65, \mu_2 = 45, \mu_3 = 25$, respectively.

Therefore, the model for the ground truth f that we used in all our simulations was

$$f = 65 \cdot 1_{\Sigma_1} + 45 \cdot 1_{\Sigma_2} + 25 \cdot 1_{\Sigma_3}.$$

Figure 2 shows f, an inhomogeneity function g, a nonnoisy corrupted image $h = f.g$, a noisy corrupted image $\tilde{h} = h + \eta$, and histograms of h and $\tilde{h}$. To save space, the figure only shows every other z-slice (we adopt this convention for all figures).

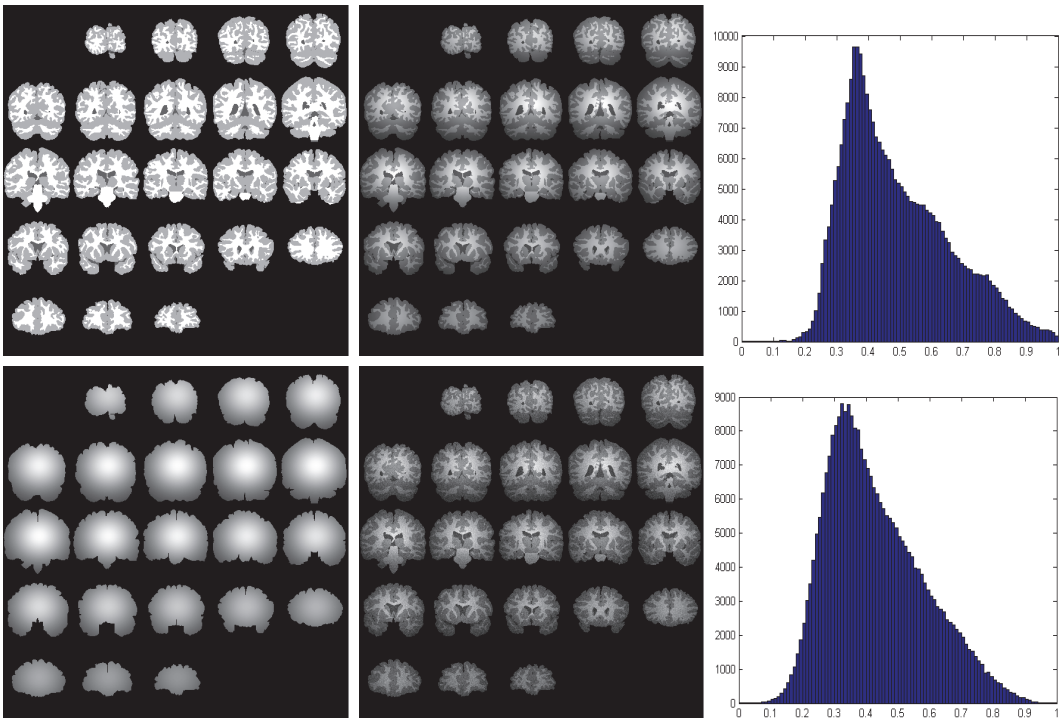


Figure 2. *Left column: ground truth signal f and true nonuniformity function g; middle column: corrupted image $h = f \cdot g$ and noisy corrupted image $\tilde{h} = h + \eta$; right column: histograms of h and $\tilde{h}$.*

For this simulated data, the true ρ values are

$$\rho_1 = \frac{65}{45} = 1.4444, \quad \text{and} \quad \rho_2 = \frac{55}{25} = 1.8.$$

8.2.1. Noise-free case. We first estimated the nonuniformity function for the noise-free case, where $h = f \cdot g$ using the true ratios $(\rho_1, \rho_2) = (1.4444, 1.8)$. We set $\beta = 1$ and $\lambda = 0.02$ and stopped the Variational Algorithm at $n = 15$.

Figure 3 shows the estimated nonuniformity function ψ and the histogram of normalized g/ψ. The normalized variance of g/ψ is 0.001 (mean = 0.9301) which compares well with the expected variance of 0 in ideal recovery.

Figure 4 also confirms our algorithm's performance using secondary performance measures. From left to right, the corrupted image h, the corrected image h/ψ, the histogram of h, and the histogram of h/ψ are shown. It is clear, visually and numerically, that ψ is accurate enough to recover the unknown regions $\Sigma_1, \Sigma_2, \Sigma_3$ as well. In fact, the estimated segmentations $\Sigma_1, \Sigma_2, \Sigma_3$ using ϕ_1, ϕ_2 were perfect. The ratios between the average values of h/ψ in the three regions are

$$\tilde{\rho}_1 = \frac{\text{average value in } \Sigma_1}{\text{average value in } \Sigma_2} = 1.4736 \quad \text{and} \quad \tilde{\rho}_2 = \frac{\text{average value in } \Sigma_2}{\text{average value in } \Sigma_3} = 1.7268,$$

which compare well with the true ρ_1, ρ_2.

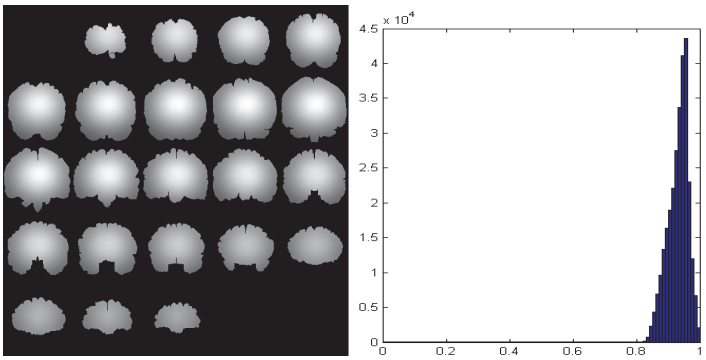


Figure 3. *Left: estimated nonuniformity ψ; right: histogram of normalized g/ψ. This has a normalized variance of* 0.001 *(mean* 0.9301*).*

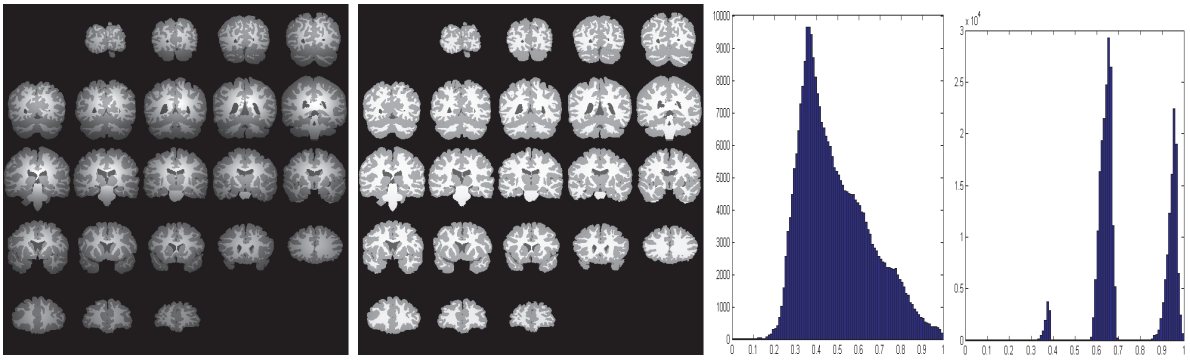


Figure 4. *From left to right: corrupted image h, corrected image h/ψ, histogram of h, and histogram of h/ψ.*

8.2.2. Noisy case. Next, we added Gaussian white noise to the corrupted image h of the previous section, to have $\tilde{h} = h + \eta$. The amount of noise added to a noise-free image was measured by the SNR (signal-to-noise ratio), which is defined by

$$SNR = 10 \times \log_{10} \left(\frac{\text{variance of noise-free image}}{\text{variance of noise}} \right) \text{ db},$$

and we used $SNR = 10$ db. Even though noise in MR images follows a Rician distribution, we consider additive Gaussian noise because the Gaussian noise model approximates the Rician noise model well when the noise level is low (SNR is relatively high), which is the case we are interested in. Hence, the focus of this section is to evaluate how well our formulation performs in estimating nonuniformity in the presence of Gaussian noise.

The result of using the algorithm with the noisy image is shown in Figure 5, where the estimated nonuniformity function $\tilde{\psi}$ and the histogram of normalized $g/\tilde{\psi}$ are displayed. The normalized variance of $g/\tilde{\psi}$ is 0.0018 (mean = 0.9150).

Figure 6 also provides the same analysis between $\tilde{h}$ and $\tilde{h}/\tilde{\psi}$ as Figure 4 from the noise-free case. The main cause for the spread in the histogram of $\tilde{h}/\tilde{\psi}$ is noise. This simulation with

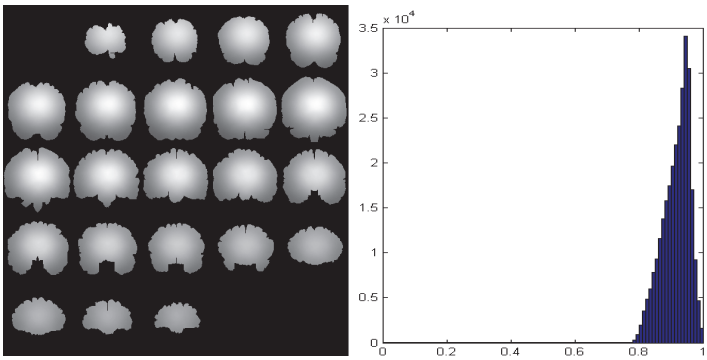


Figure 5. *Left: estimated nonuniformity $\tilde{\psi}$ for noisy data; right: histogram of normalized $g/\tilde{\psi}$. The normalized variance of g/ψ is* 0.0018 *(mean=* 0.9150*).*

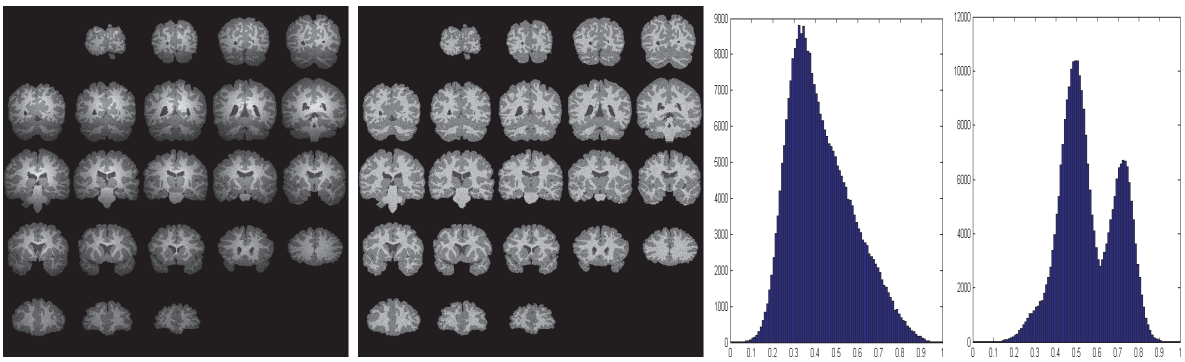


Figure 6. *From left to right: noisy corrupted image $\tilde{h}$, corrected image $\tilde{h}/\tilde{\psi}$, histogram of $\tilde{h}$, and histogram of $\tilde{h}/\tilde{\psi}$. Nonuniformity together with the unknown regions was measured well. The source of the spread in the histogram of $\tilde{h}/\tilde{\psi}$ is noise.*

noise clearly shows the robustness of our algorithm for estimating the nonuniformity in the presence of noise.

Not only was our algorithm able to correctly measure the nonuniformity, but also it estimated the unknown regions $\Sigma_1, \Sigma_2, \Sigma_3$ reasonably well. Even though our emphasis is not on the segmentation and we don't expect the segmentations to be as accurate as in the noise-free case, Figure 7 shows that the estimates are quite reasonable in this noisy case as well. To be more precise, in this simulation the differences between the true and the computed regions for Σ_1 and Σ_2 were less than 1%. As for Σ_3, the difference was about 50%, which is mainly because the true Σ_3, representing the CSF, is small in size and a slight modification to the region by a type of speckle noise seen in the figure, yielded this difference. To be consistent with the analysis from the previous noise-free case, we provide the ratios between the average values of $\tilde{h}/\tilde{\psi}$ in the three regions:

$$\tilde{\rho}_1 = \frac{\text{average value in } \Sigma_1}{\text{average value in } \Sigma_2} = 1.4783 \quad \text{and} \quad \tilde{\rho}_2 = \frac{\text{average value in } \Sigma_2}{\text{average value in } \Sigma_3} = 1.7066.$$

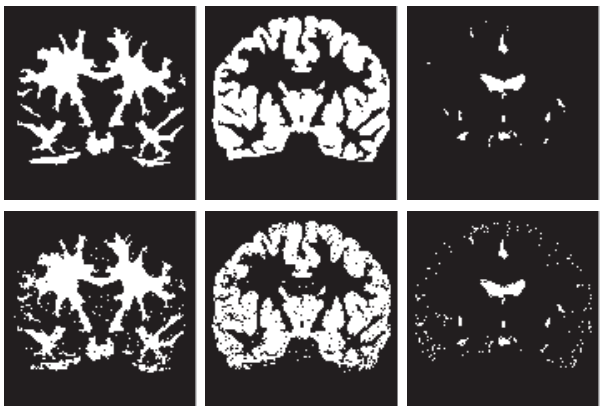


Figure 7. *One slice was chosen for comparison between the true and the estimated regions in the noisy case. Top row: the true regions $\Sigma_1, \Sigma_2, \Sigma_3$ in the chosen slice; bottom row: corresponding estimated regions.*

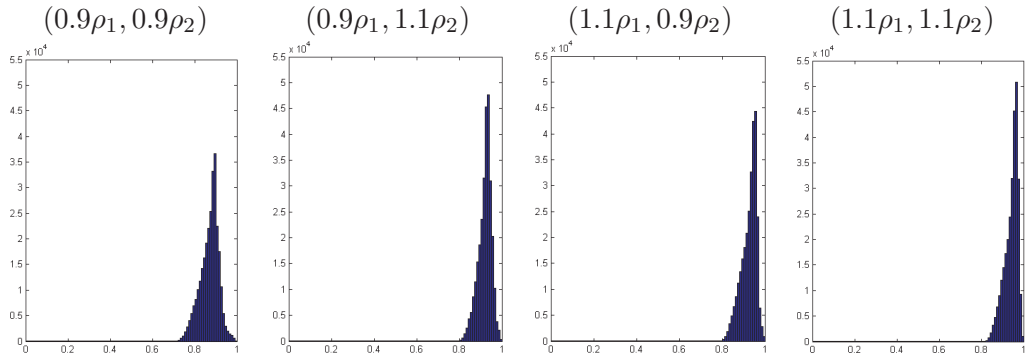


Figure 8. *Histograms of normalized $g/\tilde{\psi}$ obtained by applying the Adaptive Variational Algorithm with two outer iterations. Each histogram corresponds to the ratios above it.*

8.2.3. The adaptive variational algorithm. Next, we explored the performance of the Adaptive Variational Algorithm with the noisy data of the previous section. The algorithm was initialized to inaccurate ρ values as

$$\tilde{\rho_1} = (1 \pm 0.1)\rho_1 \text{ and } \tilde{\rho_2} = (1 \pm 0.1)\rho_2.$$

We kept the same parameters from the previous experiments except for β, which is now set to $\beta = 0.5$. This allowed the nonuniformity function $\tilde{\psi}$ to have greater flexibility in finding the correct shape. We first observed that the Variational Algorithm itself with these inaccurate ratios produced satisfactory estimations implying the algorithm is somewhat insensitive to inaccurate ratios. Moreover, the Adaptive Variational Algorithm produced even better estimates, as was expected.

We omit displaying the estimated $\tilde{\psi}$ for these cases, and instead directly display in Figure 8 the histograms of the estimated normalized $g/\tilde{\psi}$, which were obtained with the Adaptive Variational Algorithm with two outer iterations. Each column in Figure 8 corresponds to

Table 1

The mean values and the variances of the histograms in Figure 8 obtained by the Adaptive Variational Algorithm.

Given ratios	$(0.9\rho_1, 0.9\rho_2)$	$(0.9\rho_1, 1.1\rho_2)$	$(1.1\rho_1, 0.9\rho_2)$	$(1.1\rho_1, 1.1\rho_2)$
Mean value	0.8691	0.9186	0.9246	0.9398
Variance	0.0019	0.0009	0.0013	0.0010

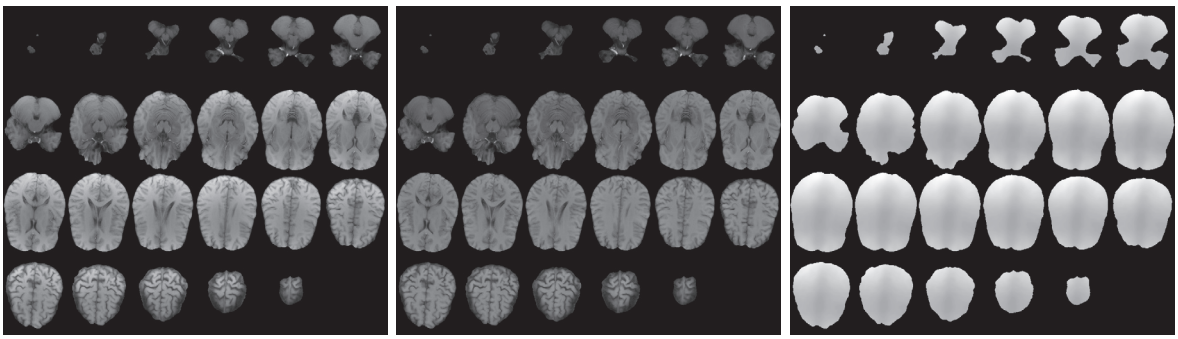


Figure 9. *Left: real MR images h of size $197 \times 154 \times 25$ provided by Magnetic Resonance Research Center, School of Medicine, Yale University; middle: corrected image h/ψ obtained with the ratios updated once; right: estimated nonuniformity function ψ. The total time elapsed for each application of the Variational Algorithm in the Adaptive Variational Algorithm was approximately 11 minutes.*

each of these four cases indicated at the top of the columns. Table 1 provides the mean values and the variances of $g/\tilde{\psi}$ shown in Figure 8.

The simulations above imply that our algorithm initialized with inaccurate ratios yields reliably good nonuniformity estimates. Moreover, the Adaptive Variational Algorithm improves the nonuniformity estimates.

As far as the degree of inaccuracy of the initial ratios is concerned, the performance becomes poor when the given ratios are off by 20% from the true ones, i.e., for $((1 \pm 0.2)\rho_1, (1 \pm 0.2)\rho_2)$. Therefore, as long as the given ratios are close to the true ones, the Adaptive Variational Algorithm will work well.

8.3. Real MR images. We have processed over 50 3 T T1-weighted brain MR images with our algorithm. The algorithm converged without any manual intervention or tweaking of parameters in all cases.

For real MR images, the Variational Algorithm turned out to be quite insensitive to the initial value of (ρ_1, ρ_2). The algorithm gave reasonable answers for ψ even at $n = 2$. With $\lambda = 0.02$ fixed, we found that $\beta \leq 2$ allows the nonuniformity function ψ to pick up local structures of the brain, and $\beta \geq 6$ makes ψ too smooth. Hence, we used $\beta = 4$. Even though the variational algorithm is not sensitive to (ρ_1, ρ_2) with real MR images, we used the Adaptive Variational Algorithm of section 7. The initial values for all images were $\rho_1 = 1.4444, \rho_2 = 1.8$.

The main difficulty in reporting the results of these images is that we do not know the true inhomogeneity g, and hence it is not possible to measure the normalized variance of g/ψ. Nevertheless, we show results from a typical case in Figure 9.

In Figure 9 the inhomogeneity seems to create brighter intensities in the frontal regions of the brain, and as the estimated $\tilde{\psi}$ shows, this has been clearly captured by the Adaptive Variational Algorithm.

8.4. Comparison results with the N3 method. We now compare the results of our algorithm with the N3 method [14] as available in the N3 package, hosted at the NITRC (Neuroimaging Informatics Tools and Resources Clearinghouse, http://www.nitrc.org). For data, we use the synthetic data sets with and without noise that are displayed in Figure 2, and the real MR image data set displayed in Figure 9.

Because the nonuniformity is estimated only up to a scale factor, we measure the performance of our method and the N3 method as follows: In simulations, where the true nonuniformity g is known, we compute the normalized histogram H of $\frac{g-\min(g)}{\max(g)-\min(g)}$ with m bins, i.e., $H : \{1, \ldots, m\} \to \mathbb{N}$. Then, we set the discrete probability P to

$$p[i] = \frac{H(i)}{\sum_{l=1}^{m} H(l)}.$$

Next, we compute the normalized histograms of the nonuniformity function estimates of our method and of the estimate from the N3 method in the same way. We denote the normalized histograms as Q_1 for our method and Q_2 for N3. We measure the closeness of Q_i, $i = 1, 2$, to P by the Kullback–Leibler distance $KL(P, Q_i)$ (KL distance) defined by

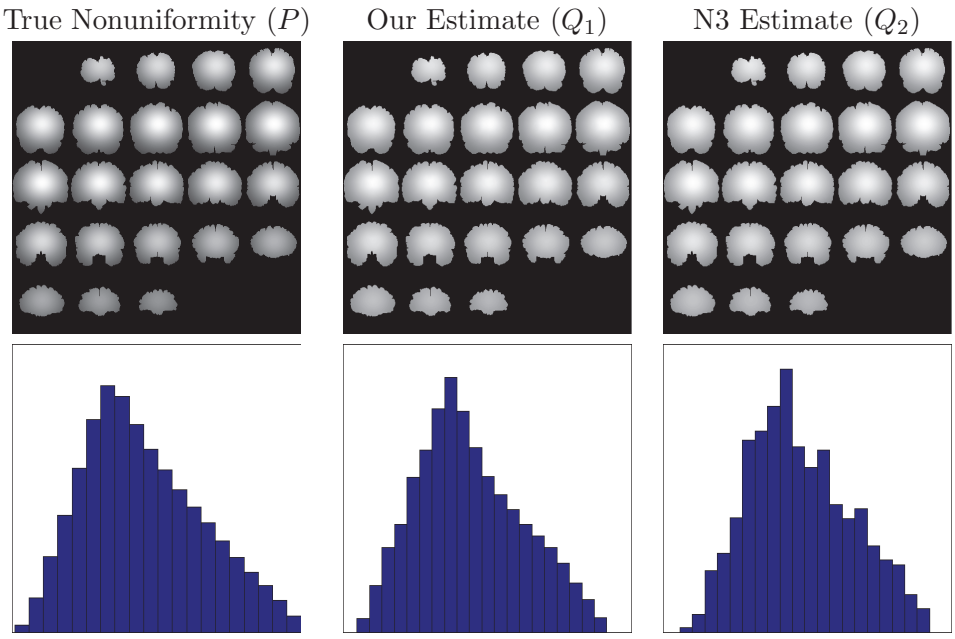
$$KL(P, Q_i) = \sum_{j=1}^{m} p[j] \ln \left(\frac{p[j]}{q_i[j]} \right),$$

where $P = \{p[1], \ldots, p[m]\}, Q_i = \{q_i[1], \ldots, q_i[m]\}$ are two discrete probability distributions. Below we report the KL distances for bin sizes, m =20, 50, and 100.

Figure 10 compares the true nonuniformity of synthetic, noise-free data with the estimated nonuniformity by our method and by the N3 method. The first row of the figure shows the true and estimated nonuniformity functions. The second row shows histograms of the nonuniformity functions using 20 bins. The KL-distance values for the histograms are displayed in a table below the second row for 20, 50, and 100 bins. For all bins, KL-distance values for our method are at least an order of magnitude smaller than for N3, clearly indicating that the nonuniformity estimate from our method is much closer to the true nonuniformity than the estimate from N3. We remark that $KL(P, Q_2) = \infty$ for m =100 bins, which happens because there are a few bins where the Q_2 histogram takes values equal to zero.

To ensure that the performance measures in Figure 10 are not sensitive to the specific nonuniformity function chosen in this example, we provide comparative evaluation of our method and N3 for another nonuniformity function in Figure 11. Here too, we can draw the same conclusion as we did through Figure 10.

Next, we test the performance of our method with that of N3 using the noisy synthetic data set in Figure 6. Figure 12 compares the histograms of the two methods just as in Figure 10. Here too the KL distances from our method are at least an order of magnitude smaller than the KL distances of N3. It is easily observed that our method is able to estimate the nonuniformity better than N3 quantitatively as well as qualitatively.

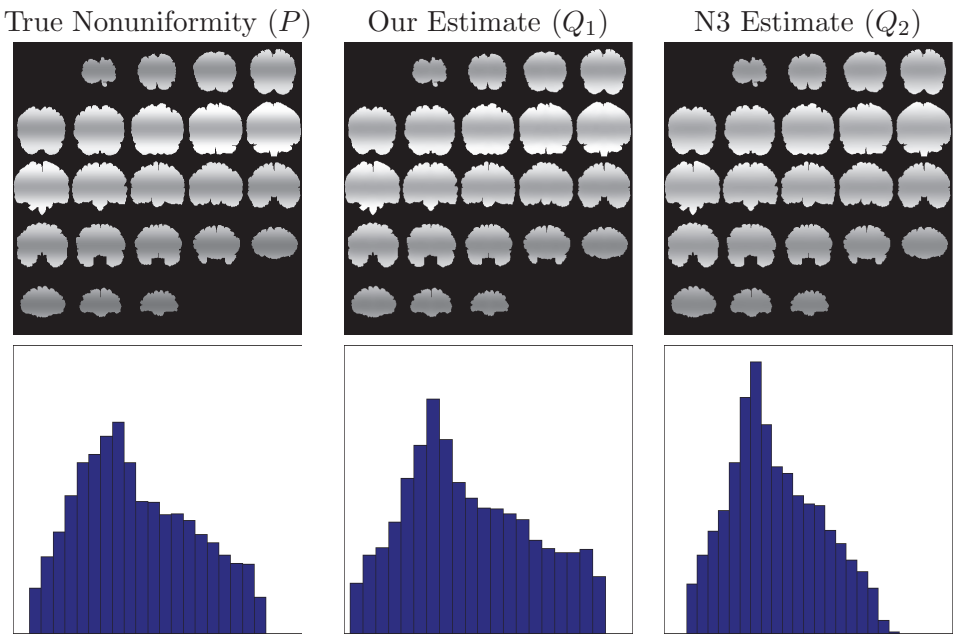


# of bins	20	50	100
$KL(P, Q_1)$	0.0023	0.0026	0.0028
$KL(P, Q_2)$	0.0240	0.0608	∞

Figure 10. *Comparing results of our algorithm with N3 using the noise-free synthetic data in Figure 4. Histograms are shown with 20 bins. Left: histogram of the true underlying nonuniformity; middle: histogram of the estimated nonuniformity by our method; right: histogram of the estimated nonuniformity by N3. The KL distance shows that Q_1 is much closer to P than Q_2 is to P.*

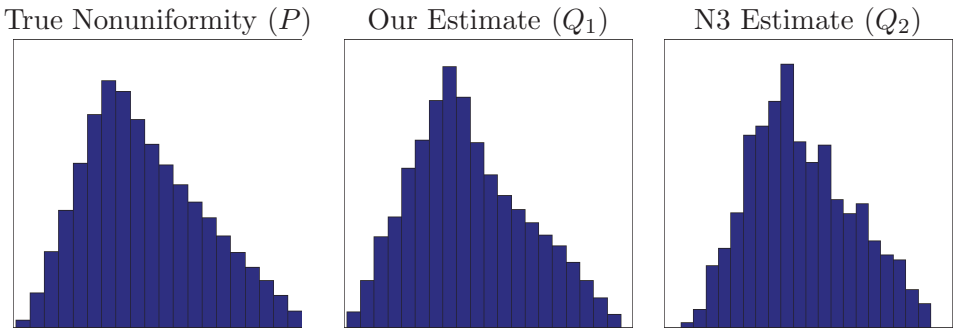
In Figure 13, we compare our method to N3 using the real MR data set shown in Figure 9. The left two subfigures are the result of our method, presented in Figure 9, and the right two subfigures are what N3 computed with the same real data set. The real nonuniformity of this data set is not available for comparing the histograms. However, the figure clearly shows that N3 is rather strict in applying its assumption that the histogram of the nonuniformity function is Gaussian.

Finally, we present the typical convergence behavior of the Variational Algorithm in Figure 14. The results presented in Figure 14 were obtained by running the algorithm with the real MR data set until steady state. The figure shows the value of the objective functional of the algorithm as a function of n, the outer iteration index in the Variational Algorithm. Each outer iteration of the algorithm contains two to three inner iterations of (4.5) followed by two to three inner iterations of (4.6). The values of the objective functional at these inner iterations are also plotted in Figure 14 in-between integer values of n. For example, there are six values of the objective functional plotted from $n = 0$ and $n = 1$. These represent the values of the objective functional at the end of three iterations of (4.5) followed by three iterations of (4.6), all of which constitute the outer iteration from $n = 0$ to $n = 1$. Figure 14 shows quite clearly that the algorithm converges rapidly in a few outer iterations.



# of bins	20	50	100
$KL(P, Q_1)$	0.0097	0.0145	0.0506
$KL(P, Q_2)$	0.0821	0.0983	0.1500

Figure 11. *Another example of comparing estimates of our algorithm to N3 with noise-free data. Histograms are shown with 20 bins. Left: histogram of the true underlying nonuniformity; middle: histogram of the estimated nonuniformity by our method; right: histogram of the estimated nonuniformity by N3. The KL distance table shows that Q_1 is much closer to P than Q_2 is to P.*



# of bins	20	50	100
$KL(P, Q_1)$	0.0055	0.0064	0.0066
$KL(P, Q_2)$	0.0240	0.0608	∞

Figure 12. *Comparison of the two methods using the noisy data set in Figure 6. The estimated nonuniformity by N3 in this noisy case appears to be very similar to that by N3 in the noise-free case.*

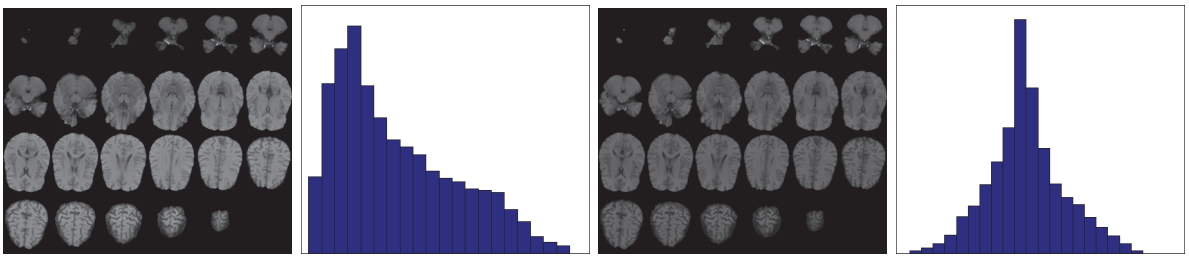


Figure 13. *Left two: the same data as the one in the middle in Figure 9 and histogram of the estimated nonuniformity by our method; right: nonuniformity corrected data and histogram of the estimated nonuniformity by N3. The fact that N3 estimates nonuniformity by using a Gaussian distribution is evident in the histogram as discussed in the introduction.*

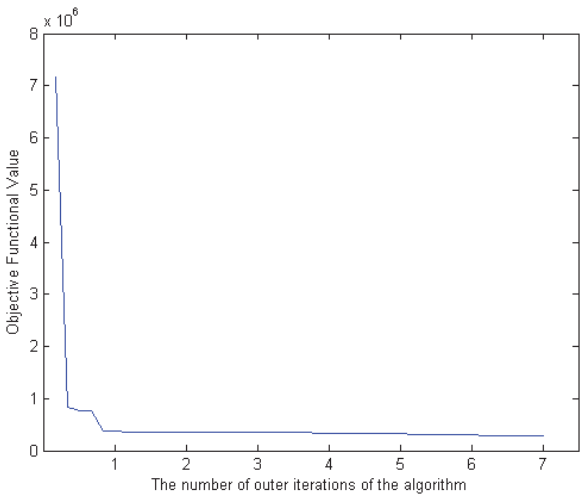


Figure 14. *Typical convergence of the algorithm. All the functional values for (4.2), (4.3), (4.4) are recorded. The x-axis represents the number of repeats of (4.5) and (4.6), which can be thought of as the total number of iterations, n, in the Variational Algorithm.*

9. Conclusion. We proposed a new formulation of the intensity nonuniformity estimation problem using variational principles. The key idea behind our formulation is this: if the true image is piecewise constant and if the ratios of the values of this image are known, then the intensity nonuniformity can be estimated by a segmentationlike variational procedure. We showed in detail how this procedure works when there are three regions in the image. Compared to other existing methods, our formulation involves only one free parameter, and its value is relatively insensitive to data. The convergence and correctness of the procedure was established rigorously. We also extended the algorithm to make it adaptive in the absence of exact prior knowledge. Experimentation with MR images showed good recovery of the inhomogeneity in 3 T T1-weighted brain images.

REFERENCES

[1] R.G. Boyes, J.L. Gunter, C. Frost, A.L. Janke, T. Yeatman, D.L. Hill, M.A. Bernstein, P.M. Thompson, M.W. Weiner, N. Schuff, G.E. Alexander, R.J. Killiany, C. DeCarli, C.R. Jack, and N.C. Fox, *ADNI study, intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils*, Neuroimage, 39 (2008), pp. 1752–1762.

[2] B. Belaroussi, J. Milles, S. Carme, Y.M. Zhu, and H. Benoit-Cattin, *Intensity non-uniformity correction in MRI: Existing methods and their validation*, Med. Image Anal., 10 (2006), pp. 234–246.

[3] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20 (2004), pp. 89–97.

[4] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[5] T. F. Chan, S. Esedoglu, and M. Nikolova, *Algorithms for finding global minimizers of image segmentation and denoising models*, SIAM J. Appl. Math., 66 (2006), pp. 1632–1648.

[6] T. Chan and L. Vese, *Active contours without edges*, IEEE Trans. Image Process., 10 (2001), pp. 266–277.

[7] C. Evans, *Partial Differential Equations*, AMS, Providence, RI, 1998.

[8] *C. Evans and R. Gariepy, Measure Theory and Fine Properties of Functions*, Stud. Adv. Math., CRC Press, Boca Raton, FL, 1992.

[9] A. Fan, W.M. Wells, J.W. Fisher, M. Cetin, S. Haker, R. Mulkern, C. Tempany, and A.S. Willsky, *A unified variational approach to denoising and bias correction in MR*, Inform. Process. Med. Imaging, 18 (2003), pp. 148–159.

[10] E. Giusti, *Minimal Surfaces and Functions of Bounded Variation*, Birkhäuser, Boston, 1984.

[11] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343.

[12] S.L. Keeling, M. Hintermüller, F. Knoll, D. Kraft, and A. Laurain, *A total variation based approach to correcting surface coil magnetic resonance images*, Appl. Math. Comput., 218 (2010), pp. 219–232.

[13] Y. Kim, *Strictly Convex Realization in Two-Phase Image Segmentation*, Technical report 13-20, UCLA CAM, Los Angeles, CA, 2013.

[14] J.G. Sled and A.P. Zijdenbos, *A nonparametric method for automatic correction of intensity nonuniformity in MRI data*, IEEE Trans. Med. Imaging., 17 (1998), pp. 87–97.