

Application of Trace-Norm and Low-Rank Matrix Decomposition for Computational Anatomy

Nematollah Batmanghelich¹

Ali Gooya¹

Stathis Kanterakis¹

Ben Taskar²

Christos Davatzikos¹

¹Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania

²Computer and Information Science, University of Pennsylvania

{nematollah.batmanghelich, ali.gooya, kanterae, christos.davatzikos}@uphs.upenn.edu
taskar@cis.upenn.edu

Abstract

We propose a generative model to distinguish normal anatomical variations from abnormal deformations given a group of images with normal and abnormal subjects. We assume that abnormal subjects share common factors which characterize the abnormality. These factors are hard to discover due to large variance of normal anatomical differences. Assuming that the deformation fields are parametrized by their stationary velocity fields, these factors constitute a low-rank subspace (abnormal space) that is corrupted by high variance normal anatomical differences. We assume that these normal anatomical variations are not correlated. We form an optimization problem and propose an efficient iterative algorithm to recover the low-rank subspace. The algorithm iterates between image registration and the decomposition steps and hence can be seen as a group-wise registration algorithm. We apply our method on synthetic and real data and discover abnormality of the population that cannot be recovered by some of the well-known matrix decompositions (e.g. Singular Value Decomposition).

1. Introduction

One of the main objectives of computational anatomy is to characterize abnormal deformation of brain anatomy. A popular method is to normalize all subjects into a standard space by registering each image with a single, universal template [8], [2]. However after normalization all normal and abnormal variations vanish from the images and get encoded in the deformation fields of the registration. Therefore, feature extraction is required to recover this information from deformation fields (e.g. determinant of Jacobian) [14, 12].

Another category of methods are generative models. One method is to represent the population by its mean. The mean can be computed in an unbiased manner [13] or can be chosen from the population [11]. A drawback of such methods is that the whole population is represented by one template and do not provide information on the subject level. Methods based on factorization of deformation fields can also be seen as a generative models. Wang et al. [19] proposed to apply Singular Value Decomposition (SVD) on initial velocity field. One can project initial velocity vectors on their principal eigen vectors however due to large variance in brain anatomies, it is unclear that the projection can help us to detect abnormality. Sabuncu et al. [17] proposed population analysis based on a mixture modeling of deformation fields (iCluster) that led to a group-wise registration. Membership value of each subject to a cluster can provide subject level information but clusters are not necessarily associated with abnormality.

The main objective of this paper is to build a generative model that describes abnormal deformations of unhealthy brains. One can consider the registration step as a generative process (let us call it P1) in which a normal template image generates normal subjects image by random deformation fields. However, the template is usually a representative of a normal ensemble and it might not be a good representation for an abnormal group. If we have another generative process (let us call it P2) to produce abnormal templates which is consistent with abnormality pattern of the abnormal group, one can cascade the P2 model to the P1 model and produce a subject that has both abnormal and normal deformations.

Our approach is to decompose the transformation (ϕ_i) that maps a template domain (Ω_T) to the i 'th subject domain (Ω_i) into the normal deformation (ψ_i) and an abnormal deformation (φ_i). We assume that abnormal deformations share common factors while normal deformations are

approximately linearly independent. To derive this decomposition, we borrowed a method proposed by Candes et al. [4] in the context of Compressed Sensing and applied it in the context of registration and population analysis. We suggest an efficient optimization algorithm that iterates between a regularized warping and the decomposition.

In the sense of population analysis, our method is similar to the iCluster [17] and can be seen as a clustering method. However our objective is to describe a population containing normal and abnormal subjects by deriving templates deformed from an unbiased normal template to describe the abnormal population.

The remainder of this paper is organized as follows: In section 2, we describe our generative setting and form a Maximum A Posterior (MAP) estimation to find the parameters of the model which leads to an optimization problem. In section 3, we will suggest an efficient algorithm to solve the optimization problem; and in section 4 the algorithm will be evaluated on synthesis dataset and real dataset consisting of normally aged and brain images with Alzheimer’s disease.

2. Method

2.1. General Setting

One of the methods to investigate image deformation is to study deformation fields that transform images to a common template as representatives of the population. Some authors suggested to apply Singular Value Decomposition (SVD) on the deformation fields to extract the most important directions of changes. However, transformations are diffeomorphic and constitute a manifold (also a Lie group \mathcal{G}), which, unlike a vector space, are not closed with respect to operations like addition. Wang et al. [19] proposed to apply SVD on initial velocity fields of the deformation as the algebra (\mathfrak{g}) of the Lie Group (\mathcal{G}). However, if the main objective of a study is to characterize abnormal deformations for population study, applying SVD directly on the velocity fields may not be helpful because normal anatomical variations have a large variance which can overshadow abnormal deformations. Hence, singular vectors are combination of normal and abnormal variations. This situation is graphically represented in Fig. 1 and an example of applying SVD on the velocity fields are discussed in Fig. 6.

Our approach is to build a generative model based on a decomposition of the transformation (ϕ_i) that maps a template domain (Ω_T) to the i ’th subject domain (Ω_i) into normal deformation (ψ_i) and abnormal deformation (φ_i). In this paper, similar to Wang et al. [19], we use velocity field to parametrize a diffeomorphism but following Arsigny et al. [1], we use stationary velocity fields to parametrize a diffeomorphism. In our notation, ϕ_i , φ_i , and ψ_i are parametrized by stationary velocity fields v_i , l_i , and s_i re-

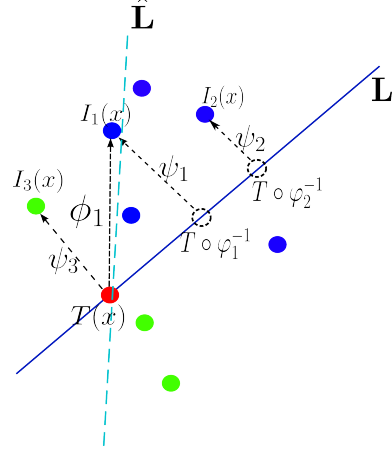


Figure 1. In this figure: T is the template and I_i ’s are images. Blue dots are abnormal and green dots are normal subjects. ϕ_i ’s are deformations that warp the template to the i ’th subject and ψ_i ’s are normal part of the deformations and φ_i ’s are abnormal part of the deformations. \mathbf{L} is true subspace of velocities of abnormal deformations and $\hat{\mathbf{L}}$ is an estimation based on velocity of ϕ_i ’s. Since normal deformations have large variance, it predominates the true estimation.

spectively.

We assume that normal anatomical differences are independent transformations (ψ_i) from identity given that the template image is unbiased mean of a normal group. Statistically speaking, we assume that given the template (T), normal subjects are identically and independently distributed.

Unlike ψ_i , we assume that abnormal deformations (φ_i ’s) are linearly correlated by common *factors*. These factors are modes of abnormal variations within abnormal ensemble. For example, in Alzheimer’s disease, hippocampus, temporal lobe, and frontal lobe endure significant shrinkage but extent of damage is different as the disease progresses. In fact, to draw an abnormal sample, the template and common factors (b_j ’s) must be known. One can simply assume that a velocity field of an abnormal deformation is a linear combination of abnormality factors (basis vectors) b_i ’s. This model is graphically represented in Fig. 2.

The main purpose of this paper is to recover the low-rank space spanned by abnormal factors b_i . One approach, for example, is Monte-Carlo method to draw samples from the graphical model shown in Fig. 2 and estimate the parameters but due to size of our problem, it is not practical. Another approach is to use Expectation Maximization (EM) to derive parameters of the model. Instead, our model is a Maximum A Posterior estimation (MAP) and it is based on a recent paper by Candes et al. [4] that is one of the variations of Robust Principal Component Analysis. Under some conditions discussed in [4], this method is able to recover

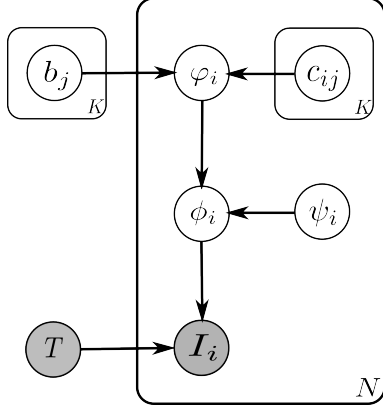


Figure 2. Graphical model representing our probabilistic model of the population. φ_i is abnormal deformation that is derived by Lie group (\mathcal{G}) *Exponential* ($Exp(\cdot)$) of l_i ($\varphi_i = Exp(l_i)$). l_i can be written as $\sum_{j=1}^K b_j c_{ij}$ where b_j are common factors (basis vectors) of abnormality and c_{ij} is loading coefficients of b_j for the i 'th subject. $\psi = Exp(s_i)$ is normal deformation and $\phi_i = Exp(v_i)$ is composition of these two transformations, namely $\phi_i = \psi_i \circ \varphi_i$; hence in our setting $v_i = l_i + s_i$. Subject images (I_i 's) and the template image (T) are given. N is number of subjects and K is number of basis vectors (common factors for abnormality).

the low-rank space and corresponding corruptions (s_i) exactly. After deriving the low-rank space, b_j and c_{ij} can be derive by any matrix decomposition (e.g. SVD).

To simplify our notation we introduce $\Phi = [\phi_1 \cdots \phi_N]$, $\mathbf{S} = [s_1 \cdots s_N]$, $\mathbf{L} = [l_1 \cdots l_N]$, and $\mathbf{V} = [v_1 \cdots v_N]$. Based on the graphical model in Fig.2, we can use MAP estimation as follows:

$$p(\Phi(\mathbf{V}), \mathbf{S}, \mathbf{L} | \{I_i\}_{i=1}^N, T) \propto p(\{I_i\}_{i=1}^N | \Phi(\mathbf{V}), T) p(\Phi(\mathbf{V}), \mathbf{S}, \mathbf{L})$$

in which $p(\cdot)$ is a posterior density function of corresponding variables. Since $\phi_i = Exp(v_i)$, we will drop the dependency on v_i to simplify our notation.

To find $(\Phi^*, \mathbf{L}^*, \mathbf{S}^*)$ that maximize the posterior probability, we can form an optimization as negative of logarithm of posterior density, namely $\mathcal{L}(\Phi, \mathbf{S}, \mathbf{L} | \{I_i\}_{i=1}^N, T) = -\log p(\Phi, \mathbf{S}, \mathbf{L} | \{I_i\}_{i=1}^N, T)$:

$$\begin{aligned} (\Phi^*, \mathbf{L}^*, \mathbf{S}^*) &= \arg \max_{\Phi, \mathbf{S}, \mathbf{L}} \mathcal{L}(\Phi, \mathbf{S}, \mathbf{L} | \{I_i\}_{i=1}^N, T) \\ &= \arg \max_{\Phi, \mathbf{S}, \mathbf{L}} [-\log p(\{I_i\}_{i=1}^N | \Phi, T) \\ &\quad - \log p(\Phi, \mathbf{S}, \mathbf{L})] \end{aligned} \quad (1a) \quad (1b)$$

Formulation of the first Eqn.(1a) comes from our noise model in Fig.2. It can be basically viewed as image registration; we call it likelihood term which will be discussed shortly. Eqn.(1b) is the decomposition step that is based on [4] and will be discussed in section 2.3.

2.2. Likelihood Term

Assuming Gaussian noise on images, namely $p(I_i | \phi_i, T) - T \sim \mathcal{N}(0, \sigma)$, we can write the likelihood term simply as sum of squared differences of error in matching:

$$\begin{aligned} \mathcal{D}(\{I_i\}_{i=1}^N, T | \Phi) &= -\log p(\{I_i\}_{i=1}^N | \Phi, T) \\ &\propto \sum_{i=1}^N (\|T \circ \phi_i^{-1} - I_i\|_2^2 + \|I_i \circ \phi_i - T\|_2^2) \end{aligned} \quad (2)$$

in which $\|\cdot\|_2^2$ is sum of squared differences over a domain. Since it is a generative model and we want to have forward ($Exp(v_i)$) and backward transformation $Exp(-v_i)$ to be valid warpings, the second term in the summand is added to symmetrize $\mathcal{D}(\cdot, \cdot)$. Similar approaches are proposed by Christensen et al. [7] and Vercauteren et al. [18].

2.3. Decomposition Term

From here on, we assume that the velocity fields (v_i 's) and corresponding decomposition terms (l_i, s_i) are discretized into voxels, hence v_i, l_i , and s_i are vectors namely $v_i, l_i, s_i \in \mathbb{R}^{dM}$ which d is dimensionality ($d = 2, 3$) and M is number of voxels of the images. We can decompose prior over ϕ_i 's, $\mathbb{P}(\Phi | T)$, as follows:

$$\begin{aligned} -\log \mathbb{P}(\Phi | T) &\propto \mathcal{R}(\Phi(\mathbf{V})) \\ &= \mathcal{R}(\Phi(\mathbf{L}, \mathbf{S})) \propto \mathcal{R}_L(\mathbf{L}) + \lambda \mathcal{R}_S(\mathbf{S}) \end{aligned} \quad (3)$$

in which $\mathcal{R}(\cdot)$'s are regularization functions for corresponding variables and λ is relative weighting between different regularizers. Each of the regularizer terms will be discussed shortly.

Having the deformation fields parametrized by stationary velocity fields, we have already assumed that the velocity fields should not be too far from 0 $\in \mathfrak{g}$. Following [4], we choose ℓ_1 -norm to encourage this property. More discussion on other choice of norm are beyond the scope of this paper but this norm works well with the regularization term we will use for \mathbf{L} :

$$\mathcal{R}_S(\mathbf{S}) = \|\mathbf{S}\|_1 \quad (4)$$

in which $\|\mathbf{S}\|_1$ is sum of the absolute values of elements of the matrix \mathbf{S} . It is known in statistical learning and optimization context [10], [5] that ℓ_1 -norm encourages sparsity which might not be the case for s_i 's, we handle this issue by iteratively solving an optimization problem which will be discussed later.

Having l_i as a linear combination of b_j 's (Fig.2) means that it belongs to a low-rank sub-space however the rank of \mathbf{L} is not known a priori. One approach is to find a decomposition ($\mathbf{V} = \mathbf{L} + \mathbf{S}$) that minimizes the rank of \mathbf{L} . It is

suggested and also natural to add the rank of \mathbf{L} as a regularization [9]. However, similar to ℓ_0 -norm, rank of a matrix has a discrete quantity and it is challenging to optimize. Recently, there has been some progress to address this issue mostly in Compressed Sensing [3, 4] and optimization context [9]. Similar to the relaxation of ℓ_0 -norm with ℓ_1 -norm ([16]), it is suggested in literature (e.g. [9]) to relax rank function with a matrix norm called *nuclear-norm* or (*trace-norm*). *Trace-norm* is similar to ℓ_1 -norm in spectral domain of a matrix. If a matrix $\mathbf{L} \in \mathbb{R}^{m \times N}$ (N is number of subjects, $m = dM$ where d is dimensionality of images and M is number of voxels of image), its nuclear norm is defined as follows:

$$\mathcal{R}_L(\mathbf{L}) = \|\mathbf{L}\|_* = \text{trace}(\sqrt{\mathbf{L}^T \mathbf{L}}) = \sum_{i=1}^{\min\{m, N\}} \sigma_i \quad (5)$$

in which $\sqrt{\mathbf{L}^T \mathbf{L}}$ denotes a matrix \mathbf{B} such that $\mathbf{B}\mathbf{B} = \mathbf{L}^T \mathbf{L}$ and σ_i 's are singular values of matrix \mathbf{L} . Note that matrix \mathbf{L} does not need to be squared matrix. Nuclear-norm encourages low-rank matrix similar to ℓ_1 -norm that encourages sparsity. Advantage of using this formulation rather than optimizing with respect to basis vectors, b_j 's, and corresponding loadings, c_{ij} 's, is two-fold: 1) it is convex with respect to parameter (\mathbf{L}) while optimizing jointly with respect to basis vectors (b_j 's) and corresponding loading coefficients (c_{ij} 's) is not a convex optimization problem, 2) number of basis vectors, r , does not need to be known and problem has one less parameter. Given Eqn.(4) and Eqn.(5) regularization on velocity field can be formulated as follows:

$$\begin{aligned} \mathcal{R}(\mathbf{V}) &= \mathcal{R}_L(\mathbf{L}) + \lambda \mathcal{R}_S(\mathbf{S}) = \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{subject to:} \quad & \mathbf{V} = \mathbf{S} + \mathbf{L} \end{aligned} \quad (6)$$

Observe that as λ decreases, Eqn.(6) favors a low rank solution (first term). Hence, one may concern about the choice of λ since it may arbitrarily bias the decomposition toward or against low-rank term. However, as it is shown by Candès et al. [4], under some mild conditions on pattern of sparsity of *true* sparse matrix (\mathbf{S}_0) and rank of *true* low-rank matrix (\mathbf{L}_0), it is enough to set $\lambda = \frac{1}{\sqrt{\max\{m, n\}}}$, then minimizers of Eqn.(6), $(\hat{\mathbf{L}}, \hat{\mathbf{S}})$, are *exact* with probability at least $1 - \beta n^{-10}$ (β is a constant). Thus, for all of the experiments of this paper, we set $\lambda = \frac{1}{\sqrt{dM}}$ which d is dimensionality ($d = 2, 3$) and M is number of pixels of the images.

Algorithm 1 General Scheme

```

1:  $k \leftarrow 0$ 
2: while not converged do
3:    $\mathbf{V}^{k+1} \leftarrow \arg \min_{\mathbf{V}} \mathcal{D}(\{I_i\}_{i=1}^N, T | \Phi(\mathbf{V}))$ 
4:    $(\mathbf{L}^{k+1}, \mathbf{S}^{k+1}) \leftarrow \arg \min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1, \text{s.t.: } \mathbf{V}^{k+1} = \mathbf{L} + \mathbf{S}$ 
5:    $I_i^{k+1} = I_i^k \circ \varphi_i^{k+1}$ , where  $\varphi_i^{k+1} = \text{Exp}(l_i^{k+1})$ 
6:    $k \leftarrow k + 1$ 
7: end while

```

3. Optimization

Assembling Eqn.(6,2), we have the following optimization problem:

$$\begin{aligned} \min_{\mathbf{V}, \mathbf{S}, \mathbf{L}} \quad & \mathcal{D}(T, \{I_i\}_{i=1}^N | \Phi) + \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{subject to: } & \mathbf{V} = \mathbf{L} + \mathbf{S} \\ & \phi_i = \text{Exp}(v_i) \end{aligned} \quad (7)$$

Our approach to optimize Eqn.(7) is a block-wise optimization method; namely we optimize with respect to blocks of parameters and iterate between blocks until a convergence criterion is met. For Eqn.(7), two sub-modules can be discerned: 1) *warping step*: optimizing with respect to $\Phi(\mathbf{V})$, 2) *decomposition step*: optimizing with respect to (\mathbf{L}, \mathbf{S}) jointly. We briefly summarize the algorithm in Algorithm 1. The warping step is almost similar to Vercauteren et al. [18] except a simple regularization. Since in the decomposition step ℓ_1 -norm is used for regularization of s_i 's which encourages sparsity and it might not be the case for normal deformation, in the first iteration, part of ψ_i may fall into φ_i . Hence, we recover s_i iteratively but as the iteration counter, k , increases, k 'th estimation of abnormal deformation (φ_i^k) should stay close to the previous estimation (φ_i^{k-1}); namely $k \rightarrow \infty, \varphi_i^k \rightarrow \varphi_i^{k-1}$. We encourage it by a penalty term.

3.1. Warping Step

In this step, the following optimization problem must be solved:

$$\begin{aligned} \min_{\mathbf{V}} \quad & \mathcal{D}(T, \{I_i^k\}_{i=1}^N | \Phi) + \gamma^k \sum_{i=1}^N \|v_i\|_2^2 \\ \text{subject to: } & \phi_i = \text{Exp}(v_i) \end{aligned} \quad (8)$$

where $\|\cdot\|_2^2$ is ℓ_2 -norm. The ℓ_2 -norm is added because the deformation (φ_i^k) is already absorbed in I_i^k and we expect new solution to be as close to $\mathbf{0} \in \mathfrak{g}$ as possible. $\gamma > 1$ is a scalar value enforcing the $v_i \rightarrow \mathbf{0}$ as $\gamma^k \rightarrow \infty$. The quadratic form does not significantly change the update rules for the velocity term proposed in [18]; namely it adds a linear term to the update rule of the backward term in [18] and forward update rule and other parts of their algorithm

stay unchanged:

$$u(p) = \frac{T(p) - I_i^k \circ \phi_i(p)}{\|J^p\|^2 + (\gamma^k + \frac{\sigma_i^2(p)}{\sigma_x^2})} J^p - \frac{2\gamma^k v_i(p)}{\|J^p\|^2 + (\frac{\sigma_i^2(p)}{\sigma_x^2} + \gamma^k)} \quad (9)$$

where $\sigma_i(p) = |T(p) - I_i^k \circ \phi_i(p)|$ is local estimate of noise in p 'th pixel of images and $J^p = -\frac{1}{2}(\nabla_p^T(I_i^k \circ \phi_i(p)) + \nabla_p^T T)$ and $u(p)$ is update for p 'th pixel of velocity field. For more detail refer to [18]. Observe that in the equation, forward transformation (ϕ_i) is regularized and since Eqn.(2) is inverse consistent, the backward transformation (ϕ_i^{-1}) is regularized implicitly by the procedure explained in [18]. Also notice that 8 can be optimized in a parallel manner.

3.2. Decomposition Step

The decomposition step involves solving the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{S}} \quad & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ \text{subject to:} \quad & \mathbf{V}^{k+1} = \mathbf{L} + \mathbf{S} \end{aligned} \quad (10)$$

This optimization problem can be reformulated as a semidefinite program (SDP) [6] and can be treated as general SDP problem and be solved by any off-the-shelf SDP solver. However, most of the SDP solvers use interior method that do not scale well for large matrices because they rely on the second-order information of the objective function. This is also the case for our problem. Computational complexity of the step direction using interior method is $O(m^6)$ that renders it computationally prohibitive for our case. To overcome the scalability issue, we need to use only the first order information and entirely harness the special properties of Eqn.(10). This problem was addressed recently by Wright et al. [20]. Authors proposed to use Augmented Lagrange Multiplier (ADM) to recover corrupted low-rank matrices. They proposed very fast first order method to solve Eqn.(10) for large matrices and they successfully applied the algorithm for background modeling in video surveillance.

Following [20], augmented Lagrangian of Eq.(10) is:

$$\begin{aligned} h(\mathbf{L}, \mathbf{S}, \mathbf{Y}, \mu) = & \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \\ & + \langle \mathbf{Y}, \mathbf{V}^{k+1} - \mathbf{L} - \mathbf{S} \rangle \\ & + \frac{\mu}{2} \|\mathbf{V}^{k+1} - \mathbf{L} - \mathbf{S}\|_F^2 \end{aligned} \quad (11)$$

where \mathbf{Y} and μ are Lagrangian multipliers and $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{tr}(\mathbf{X}\mathbf{Y}^T)$ is trace of matrix multiplication between \mathbf{X} and \mathbf{Y} (an inner product in a matrix space) and $\|\cdot\|_F$ is Frobenius norm.

The idea to optimize Eqn.(10) is as follows: for a given Lagrange multipliers (\mathbf{Y}, μ), unconstrained Lagrangian (Eqn. (11)) has a closed form solution for primal variables (\mathbf{L}, \mathbf{S}) using singular value decomposition (SVD)

Algorithm 2 Estimating low-rank and sparse velocity fields: ($\mathbf{L}^*, \mathbf{S}^*$)

```

1:  $\mathbf{Y}_0 = \mathbf{Y} / \max(\|\mathbf{V}^{k+1}\|_2, \|\mathbf{V}^{k+1}\|_\infty)$ 
2:  $\mathbf{S}_0 = \mathbf{0}; \mu_0 > 0; \rho > 1; k \leftarrow 0$ 
3: while not converged do
4:    $(\mathbf{U}, \mathbf{\Sigma}, \mathbf{W}) \leftarrow \text{svd}(\mathbf{V}^k - \mathbf{S}_k + \mu_k^{-1} \mathbf{Y}_k)$ 
5:    $\mathbf{L}_{k+1} \leftarrow \mathbf{U} \mathcal{F}_{\mu_k^{-1}}[\mathbf{\Sigma}] \mathbf{W}^T$ 
6:    $\mathbf{S}_{k+1} \leftarrow \mathcal{F}_{\mu_k^{-1}}[\mathbf{V}^{k+1} - \mathbf{L}_{k+1} + \mu_k^{-1} \mathbf{Y}_k]$ 
7:    $\mathbf{Y}_{k+1} \leftarrow \mathbf{Y}_k + \mu_k(\mathbf{V}^{k+1} - \mathbf{L}_{k+1} - \mathbf{S}_{k+1})$ 
8:    $\mu_{k+1} \leftarrow \rho \mu_k$ 
9:    $k \leftarrow k + 1$ 
10: end while
```

and soft-thresholding (Eqn.(12)); however, Lagrange multiplier should be chosen such that the equality constraint in Eqn.(10) is satisfied; this leads to simple update rules for \mathbf{Y} and μ . We iterate between these two steps until a convergence criterion is satisfied. To achieve even more efficient method, we used an Inexact Augmented Lagrangian Multiplier (IALM) proposed and explained in [15] that requires one SVD steps per iteration which makes the algorithm very efficient for large scale problems like ours. Due to space limitation, detail of the algorithm is omitted and only general scheme is provided in Algorithm2 in which $\text{svd}(\cdot)$ is singular value decomposition and $\mathcal{F}_\epsilon(\cdot)$ is soft-thresholding function namely:

$$\mathcal{F}_\epsilon[x] = \begin{cases} x - \epsilon, & \text{if } x > \epsilon \\ x + \epsilon, & \text{if } x < -\epsilon \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

For more detail on Algorithm2 refer to [15].

4. Experiments

In order to evaluate the algorithm, two experiments are conducted. In the first experiment, simple simulated images are created that satisfied the assumptions of the algorithm. We will investigate how successfully the algorithm can decompose the deformation into normal and abnormal deformations and factors out the normal deformations which are presumably due to normal variations between subjects. In the second experiment, we apply our method on real anatomical MRI images of two groups of brains: Alzheimer's disease (AD) and normal control (NC). To assess the performance of the method, we measure how successfully we can differentiate between the groups with a simple algorithm. In addition, we visualize group difference by applying traditional voxel-based analysis and compare areas appearing as group difference with those that have been reported in the medical literature.

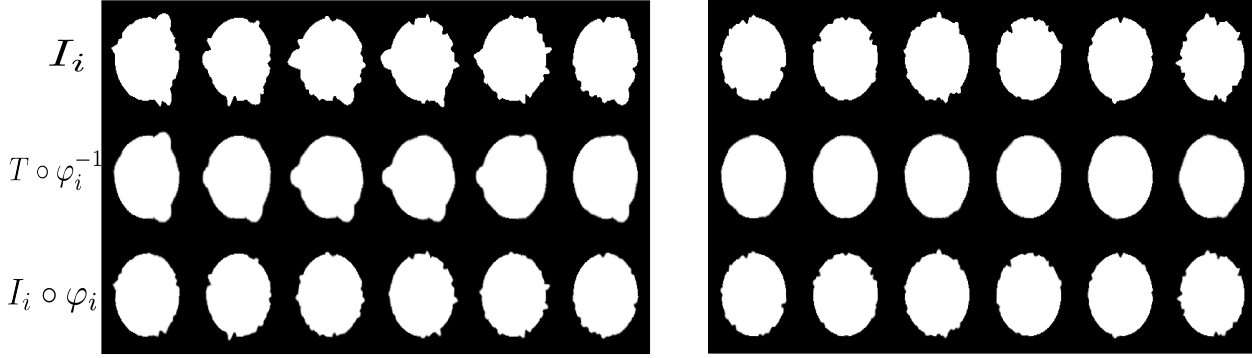


Figure 3. Results of our decomposition method on simulated data (Right: *Normal*, Left: *Abnormal*). First Row: Original images (I_i), The second row: Forward transformation of the template (T) with only abnormal part (\mathbf{L}) of the velocity field ($T \circ \varphi_i^{-1}$); the third row: backward transformations of the original images (I_i 's) with abnormal part of velocity fields ($I_i \circ \varphi_i$). In fact φ_i compensates for abnormal deformation in i 'th and $I_i \circ \varphi_i$ shows how i 'th subject would have looked like if it had not had abnormality.

4.1. Simulated Data

Simulated images are divided into two groups: Normal and Abnormal each consisting of 200 images. Normal images created by perturbing inward/outward the periphery of an optimal template (T) that is a circle. This type of deformations can happen in any place on the circle with random magnitudes and they are considered as normal deformations because they are not correlated (first row of right hand side of Fig.3 shows some examples). Abnormal images undergo two types of deformations, normal deformations which are simulated in the same way as the normal group and abnormal deformations that are correlated through three common factors or basis vectors (b_j 's in Fig.2). These basis vectors resemble abnormal bulges in three places on the periphery of the template (circle). For each abnormal subject, locations on periphery of the template (2 out of 3) and magnitude of bulging are chosen randomly (some examples are shown in the first row of Fig.3).

In this experiment, circle is used as the template that is unbiased estimation of the normal population. The second row of Fig.3 shows forward transformation of the template (T) with only abnormal part (\mathbf{L}) of the velocity fields, $T \circ \varphi_i^{-1}$. We expect φ_i to capture perturbations of i 'th subject up to the normal deformation. Since there is no abnormal deformations in a normal case, $T \circ \varphi_i^{-1}$ is approximately original template T . For an abnormal case, there exists abnormal deformation that can not be described by ψ_i 's. The third row of Fig.3 shows backward transformations of subjects. Due to absence of abnormality in the normal ensemble, $I_i \circ \varphi_i$ is approximately I_i but for an abnormal case backward transformation compensates abnormal deformation and shows how a subject, I_i , looks like before abnormal deformation.

Notice that the algorithm is not aware of the class labels (which subject is normal or abnormal) nevertheless it is able to collect linearly correlated velocity fields into \mathbf{L} matrix successfully. We also tried small abnormal deformations that are not visually significant and the algorithm reasonably recovers abnormalities. However, large abnormal deformations are shown in Fig.3 for the sake of illustration.

4.2. Real Data

We applied the algorithm on a dataset of brain MR-images consisting of two groups: 43 normally aged brains as normal controls (NC) and 40 brain images diagnosed with Alzheimer's disease (AD). All images are skull-removed and linearly aligned (affine registration) to a template to eliminate major rotations or displacements. We used the method proposed by Avants et al. [2] on the normal ensemble to build an unbiased template of the normal group.

Fig.4 shows two examples from the AD group and one example from the NC group and application of corresponding abnormal deformation fields on the template image. For a normal case, we expect almost no abnormal deformation hence $T \approx T \circ \varphi_i^{-1}$ because we have assumed that normal cases are not correlated and consequently corresponding l_i 's are small or zero. For a AD case, $T \circ \varphi_i^{-1}$ presumably shows corresponding abnormal deformation on the template and $I_i \circ \varphi_i$ presumably compensates for abnormal deformation but keeps normal anatomical differences intact.

Traditionally in voxel-based analysis, t-test is applied on features extracted from deformation field (e.g. determinant of Jacobian of a deformation field) to extract local expansion or shrinkage of tissue [12]. Although it is possible to apply similar method for this dataset or combine a feature

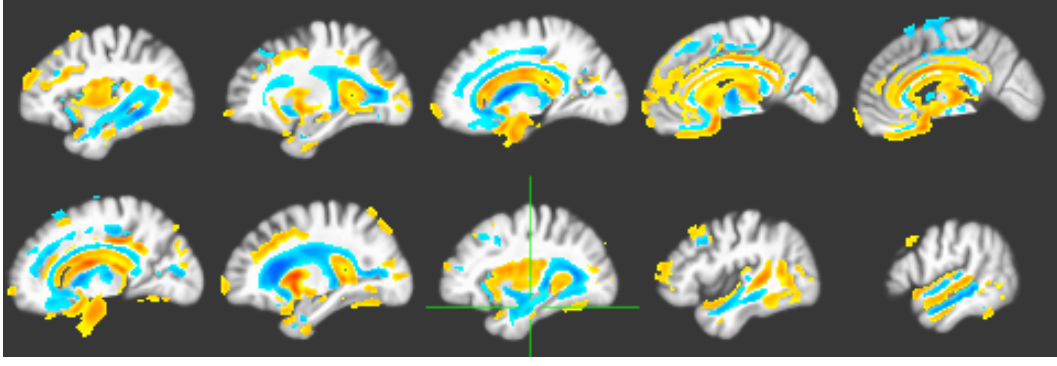


Figure 5. Applying t-test between AD and NC group: the gray matter of the template is deformed by abnormal part of the deformations ($T_{GM} \circ \varphi_i^{-1}$). The figure shows areas of statistically significant difference ($p < 0.005$) on different sagittal cuts. Underlay image is the anatomical template and the overlay image is t-value shown for $p < 0.005$.

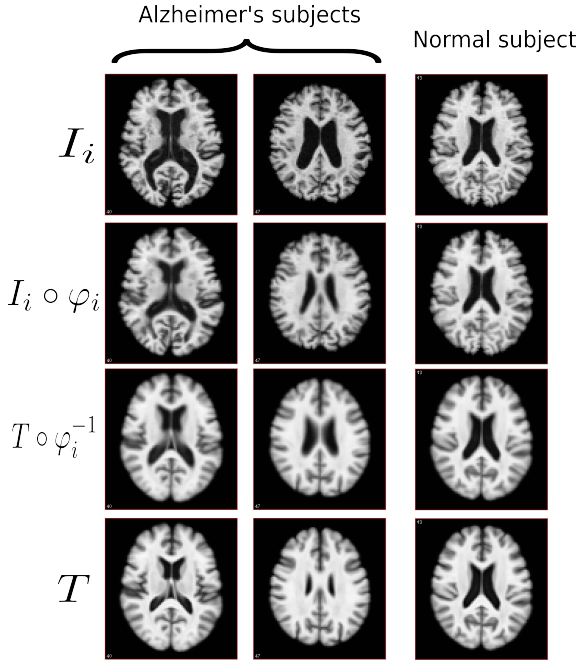


Figure 4. Two examples of Alzheimer's cases and one example of normal cases are shown. The first row shows a slice of original images. The second row shows application of φ_i on I_i that compensate abnormality in the subjects (similar to Fig.3). In fact, it shows how the cases would have looked like, if they had not had abnormality. In normal cases, there is almost no change. The third row shows deformation of the template with abnormal deformation (similar to the second row in Fig.3). the last row is original unbiased template in the same slice. In the normal case $T \approx T \circ \varphi_i^{-1}$ but in the abnormal cases, they show corresponding deformations of the subjects on the template image. Notice changes in ventricle size.

extraction scheme with our method, we keep the feature extraction as simple as possible and simply apply t-test be-

tween groups on the template deformed by the abnormal deformations of each subject $T \circ \varphi_i^{-1}$. If the generative model extracted abnormal information in φ_i 's, we should be able to find areas of difference that is consistent with medical literature only by comparing $T \circ \varphi_i^{-1}$ between two groups. Fig.5, shows areas of difference that are statistically significant ($p < 0.005$) for gray matter (GM); for example Hippocampus is revealed as one of the areas undergoing significant shrinkage between the NC and the AD groups. It is also frequently reported in medical literature. Notice that no explicit information from deformation field (e.g. determinant of Jacobian) is used, nevertheless $T \circ \varphi_i^{-1}$'s suffices to detect the group difference.

We also evaluated our method in term of classification rate. We applied simple linear Support Vector Machine (SVM) classifier with ℓ_1 -norm regularization on gray matter of $T \circ \varphi_i^{-1}$; although extracted features are simply binary image of GM mask of the template under abnormal deformation, 84.3% of subjects are classified correctly in 10-fold cross validation. Similar operation were applied on white matter (WM) and ventricle (VN) tissue types and classification rates were 81.9% and 81.9% respectively. Notice that only simple binary tissue types (for GM, WM, or VN) are used and no complicated feature extraction is applied to improve the classification rate; yet it is possible to boost the performance via other features (e.g. Jacobian of deformation field).

In order to compare singular vectors of \mathbf{L} with those of original velocity fields, we plotted accumulated sum of singular values of \mathbf{L} and \mathbf{V} for 83 eigen vectors. Since number of subjects are less than the dimensionality, number of non-zero singular values are at most number of subjects which is 83. Fig.6 shows that effective rank of \mathbf{L} is almost 40 which is equal to the number of AD subjects but normalized accumulated sum of singular values of \mathbf{V} grows almost linearly meaning that adding a new subject almost adds new rank to

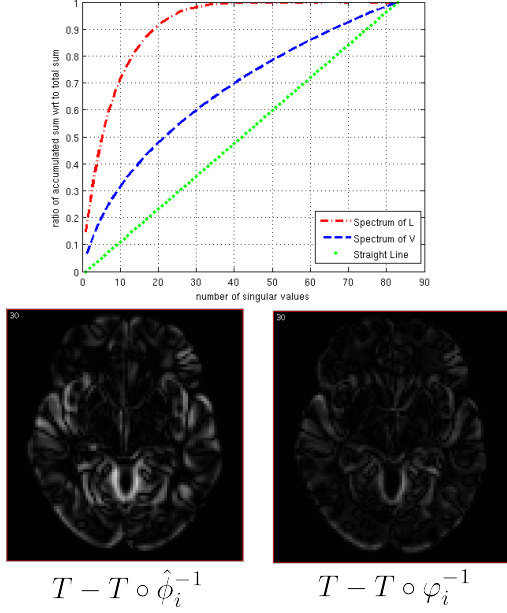


Figure 6. This figure compares compactness of singular values of \mathbf{L} with that of \mathbf{V} . The red dashed-line shows accumulated sum of singular values for \mathbf{L} normalized with $\|\mathbf{L}\|_*$ and the blue line shows the same quantity for \mathbf{V} . Notice that although the algorithm is not aware of subject labels (AD/NC), the effective rank of \mathbf{L} is almost number of AD subjects (40 subjects). The second row shows that φ_i does not deform the template for a normal subject as expected but an estimation based on \mathbf{L} with similar rank as \mathbf{L} does not serve this purpose.

\mathbf{V} . We also visually compare φ_i and reconstruction of $\hat{\phi}_i$ (we call it $\hat{\phi}_i$) on a normal subject. The number of singular vectors used to reconstruct $\hat{\phi}_i$ is the same as effective rank of \mathbf{L} . While φ_i keeps the template almost unchanged for a normal subject (as expected), $\hat{\phi}_i$ tries to make the template similar to the subject. This visual comparison shows the unique role of ℓ_1 -norm in Eqn.10 in modeling normal deformation.

5. Discussion

In this paper, we assumed that the velocity fields of normal subjects have high variance but uncorrelated while abnormal subjects are correlated through common factors (low-rank subspace) of abnormality. We proposed a decomposition that is able to discover this subspace from such high variance. It might not be very realistic to assume that there is no subspace shared between subjects of the normal group. This idea will be addressed in the future by introducing a low-rank term for normal ensemble and adding smarter priors for abnormal group to confine search space of the low-rank term for the abnormal ensemble. However, even such simple assumption as ours can recover the subspace of ab-

normality relatively well.

We also aware that elements of time invariant velocity fields cannot generate all elements of the Lie group \mathcal{G} but numerical results show that it is a reasonable approximation for our application plus it keeps our computation efficient and tractable.

References

- [1] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache, *A log-euclidean framework for statistics on diffeomorphisms*, Proc. MICCAI06. 2
- [2] B. Avants, P. Yushkevich, J. Pluta, and J. C. Gee, *The optimal template effect in studies of hippocampus in diseased populations*, Neuroimage. 1, 6
- [3] J. Cai, E. Candes, and Z. Shen, *A singular value thresholding algorithm for matrix completion*, SIAM Journal on Optimization. 4
- [4] E. Candes, X. Li, Y. Mi, and J. Wright, *Robust principal component analysis?*, preprint. 2, 3, 4
- [5] E. Candes and T. Tao, *The dantzig selector: Statistical estimation when p is much larger than n* , Ann. of Statistics 6 (2007), no. 35, 2313–2351. 3
- [6] V. Chandrasekharan, S. Sanghavi, P. Parillo, and A. Wilsky, *Rank-sparsity incoherence for matrix decomposition*. 5
- [7] G.E. Christensen and H.J. Johnson, *Consistent image registration*, IEEE Trans. Med. Imag. 7 (2001), no. 20, 568–582. 3
- [8] T.M. Peters D.L. Collins, P. Neelin and A.C. Evans, *Automatic 3d intersubject registration of mr volumetric data in standardized talairach space.*, J. of Comp. Ass. Tomog. 2 (1994), no. 18, 192–205. 1
- [9] M. Fazel, H. Hindi, and S. Boyd, *Rank minimization and applications in system theory*, Proc. American Control Conference. 4
- [10] J. Graca, K. Ganchev, B. Taskar, and F. Pereira, *Posterior vs. parameter sparsity in latent variable models*, NIPS’09. 3
- [11] A. Guimond, F. Meunier, and J. Thirion, *Average brain models: A convergence study*, Technicel Report 3731 INRIA. 1
- [12] R. Frackowiak I. Johnsrude C. Price J. Ashburner, C. Hutton and K. Friston, *Identifying global anatomical differences: Deformation-based morphometry*, Human Brain Mapping. 1, 6
- [13] S. Joshi, B. Davis, M. Jomier, and G. Gerig, *Unbiased diffeomorphic atlas construction for computational anatomy*, NeuroImage. 1
- [14] N. Lepore, C. A. Brun, A. W. Toga, J.T. Becker, and P. M. Thompson, *Multivariate statistics of the jacobian matrices in tensor based morphometry and their application to hiv/aids*, MICCAI’06. 1
- [15] Z. Lin, M. Chen, L. Wu, and Y. Ma, *The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices*, Submitted to Mathematical Programming (UIUC Technical Report UILU-ENG-09-2215). 5
- [16] A. Y. Ng, *Feature selection, l_1 vs. l_2 regularization, and rotational invariance*, Proc. ICML’04. 4
- [17] M.R. Sabuncu, S.K. Balcı, M.E. Shenton, and P. Golland, *Image-driven population analysis through mixture modeling*, IEEE Trans Med Imaging. 9 (2009), no. 28. 1, 2
- [18] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, *Symmetric log-domain diffeomorphic registration: A demons-based approach*, Proc. MICCAI08. 3, 4, 5
- [19] L. Wang, F. Beg, and M.I. Miller, *Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type.*, IEEE Trans Med Imaging 4 (2007), no. 26. 1, 2
- [20] J. Wright, A. Ganesh, Y. Peng S. Rao, and Yi Ma, *Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization*. 5