Summary of BRAIN Research Ideas. Intro here, why it's important to study the brain.

Subspace Clustering Modeling data as a union of subspaces is a simple and flexible model that fits well to the context of what are our data again? Algorithmically, the problem of estimating the union of subspaces from data defines the problem of *Subspace Clustering*, or clustering vectors into groups that lie in or near the same subspace. A union of k r-dimensional subspaces itself spans a subspace which has dimension at most kr. However, the union of subspaces model is often more flexible than the kr-dimensional subspace. Consider the illustration in Figure 1; though the two lines (two one-dimensional subspaces) are a simpler model than the two-dimensional plane, the set they define is non-convex.

Let X be a what size data matrix, where the rows represent what and the columns what. The goal of subspace clustering is to infer the underlying subspaces from X and to cluster the columns of X according to the subspaces. This would in turn give us an understanding of potential clusters of connectivity causing activity in the brain, cite here. From our preliminary results, subspace clustering is a much better tool for clustering activity regions than spectral clustering, which is the state-of-the-art approach cite. There are three open problems that we wish to address.

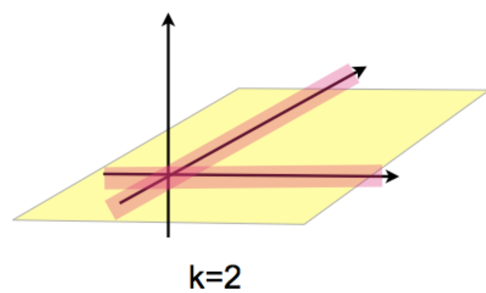

Figure 1: Illustration of two 1-d subspaces and one 2-d subspace; though the two pink lines (two one-dimensional subspaces) are a simpler model than the yellow two-dimensional plane, the set they define is non-convex.

1. Adapting the brain region clustering problem to the framework of subspace clustering, and studying the similarities and differences to the spectral clustering approach.
2. Analyzing the subspace clustering algorithm in this context, which we expect to result in both algorithmic and statistical convergence guarantees.
3. Something else here, like perhaps dealing realities of actual brain data, applying it, maybe missing data and corruptions, etc.

1 Spectral Clustering versus Subspace Clustering

This is our first theoretical contribution. The connections are clear in certain problems, give example.

2 Subspace Clustering Analysis

Understanding the theoretical properties of the subspace clustering algorithm is crucial to creating better statistical tests for brain connectivity. The current state-of-the-art algorithms rely heavily on more here and...

Algorithmic Analysis The subspace clustering problem is non-convex and finding an optimal solution to the subspace clustering problem is NP-hard [?]. However, it is believed that under mild conditions separating the subspaces, properly initialized simple algorithms like the EM algorithm should provably succeed in identifying the true subspaces underlying the data. A second theoretical contribution of our work will be to pursue this theorem of algorithmic global convergence.

Laura will talk here about avenues we could pursue.

Statistical Analysis Consistency, of course.

3 Missing-Data Subspace Clustering (or general aspects of the reality of data)

Low-rank matrix completion allows one to estimate a single subspace from incomplete data, and this work has recently been extended for the subspace clustering problem [?, ?, ?]. However, the algorithm analyzed in [?] is computationally demanding and intractable in the number of clusters. While the algorithms in [?, ?] are computationally efficient, they have no convergence guarantees.