# Sparse Methods for Medical Imaging Decomposition–sketchpad

Ben Kandel

June 10, 2012

**Abstract**

This is the abstract of the document. In short, it's clear to me that sparsity is cool and worthwhile. Now what to do with it...

I have no idea what to do with it. Oh well. We'll have to figure it out.

## 1   Methodology

There are a lot of things.

1. One basic idea is to do some **non-linear** sparse dimensionality reduction–not CCA, not PCA, but sparse non-linear dimensionality reduction. There are about a bajillion non-linear dimensionality reduction techniques [14]. Rueckert and his gang are using locally linear embedding, a manifold learning technique, for combining multimodal (in this case, imaging + clinical data) for classifying images [17]. Possibly sparse extensions of this (in his case, he used a kind of sparse representation of the image–only looked at the MTL/hippocampus, but it may be interesting to look at extensions of this). One thing to look at is sparse (in the INPUT space [12], not the feature space) kernel PCA or CCA. I don't think that anyone's done this yet for imaging data, and it might be a good place to start.

2. Contiguous PCA–in practice, NMF does, in theory [5] give a parts-based representation, when the population you're looking at actually comes from bases and loading vectors. How this applies to real brains is anyone's guess. I think that by actually enforcing contiguity constraints, you might be able to do a better job than just hoping for contiguity by doing the l1 norm or something like that.

3. Isomap [13] and LLE [10] (or robustified [11], and/or supervised, [9, 18]) are both nonlinear dimensionality reduction techniques. I wonder if they can be extended to have a sparse representation of images. The way this would work is that you'd get some sort of sparse bases of the images and then calculate graph distances and build graphs. Ideally, there would be some sort of feedback so that the classification informs the dimensionality reduction so that the reduction/sparsification is optimally tuned to produce a graph that's useful for classification. [6] makes an interesting point–it (and others) point out that Isomap might not be so great for classification. They suggest using a weight that makes images of the same class be considered "similar", or "closer", to each other if they're of the same class and further away if they're part of different classes. That's an interesting idea and I kind of like it–can it be extended to a situation in which not only the classes are known, but also the sparsity? As in, is there any way to include the sparsification directly into the weighting function? That might be useful.

   Just as a quick thought–in LLE, it works by computing the distances from each image to each other image, which gives distances, then optimizes weights, then optimizes coordinates (or something like that). In a basic sense, the "distance" metric would be replaced. That is, the distance metric is written as

$$\sum_i (I - J)^2 \tag{1}$$

   where $I$ and $J$ are the two images and the pixels are indexed by $i$. So that's a norm. So I'd rewrite this as

$$\|I - J\|, \tag{2}$$

   same notation. Using this, I would then look instead for the vector that maximizes reconstruction accuracy based on the surrounding points, but with the constraint that the vector is sparse and/or contiguous. So that would be something like

4. Semi-supervised learning–the way Kayhan does it [2, 1], he makes the labels of similar images be similar. It might instead be possible to maximize the margin between the different classes. Not sure if this would change all that much, but it might . . .

5. Sparse versions of kernel PCA and CCA are, as far as I can tell, only intermittently pursued

6. Group lasso [19], using anatomical priors? . . . This looks like a pretty obvious thing to do. Really obvious, I would say. Unbelievably obvious. Mind-blowingly obvious.

7. Christos [7] used SVM's to classify brains (AD vs. not). In particular, he makes the point that nonlinear classifiers are *fundamentally* non-interpretable, because they rely on weird combinations of data (like, say, size of hippocampus + how curvy it is + whatever), and it's not easy to represent that visually. Nevertheless, I think that it still may be possible to have some sort of sparsity in nonlinear modeling.

8. The idea of taking sparse support for SVM's is not new in AD classification. See, for example, [15]. But in that study, they had a very ad-hoc way of calculating important voxels and spatial information–they just looked at which voxels had the highest weights, and then included some voxels around them. This approach in particular seems like a dead ringer for a fused-lasso type approach. The classical reference for image classification using SVM's is [3], but there, they used histograms, not actual images (because of the problems of high-dimensionality). Again, it seems to me that regularizing based on sparsity and contiguity may help this problem. In terms of regularization, see [4]–they do real spatial regularization!!! I think that that is really not so far from what I want; I just want to jiggle things a smidgen differently.

9. I also want to emphasize that I don't want to assume non-negativity. Why? Because by assuming negativity, I can overcome bias enforced by sparsity. Viz., if you enforce sparsity, you will normally drive entries with very low numbers to 0. But that's not really what you want–you may really be interested in areas of the brain that have much lower densities/FA's/cortical thickness/whatever than other parts of the brain! So what probably makes more sense is to just to a warp to a standard template space, and then get the difference between each individual brain and the template brain, and then do decompositions on *that* space. If you don't do this, and just assume non-negativity, and use raw images, what you're going to be getting is the few areas in the brain that have the *most* activity, with every other part of the brain, because of sparsity, be sort of left out of the model (because if you don't reconstruct it, it's 0 anyway), and so won't contribute to the model. But that doesn't make sense! Because you really *are* interested in knowing whether a specific region of the brain is bedavka of low FA. In other words, if you don't have the regions of low density *explicitly* represented in a specific vector, you don't ever know how it's correlated with your outcome/label.

papers i'm reading: Manifold elastic net for sparse learning; Face Recognition Using Sparse Representations and Manifold Learning; Sparse ICA via cluster-wise PCA; Extensions of sparse canonical correlation analysis with applications to genomic data [16]; Large-Scale Sparsified Manifold Regularization; Linear Manifold Regularization for Large Scale Semi-supervised Learning (related); `http://www.iro.umontreal.ca/~lisa/workshop2004/slides/ouimet_slides.pdf`; Similarity-Based Sparse Feature Extraction Using Local Manifold Learning (might be a good idea); Laplacian Eigenmaps for Dimensionality Reduction and Data Representation (this should be a citation); Regaining sparsity in kernel principal components; Sparse Kernel Canonical Correlation Analysis; Nonlinear denoising and analysis of neuroimages with kernel principal component analysis and pre-image estimation; `http://www.cs.columbia.edu/~kathy/cs4701/documents/jason_svm_tutorial.pdf`; Sparse SVM's: `http://jmlr.csail.mit.edu/papers/volume3/bi03a/bi03a.pdf`, Dimensionality Reduction via Sparse Support Vector Machines–this looks like not such a great paper, but definitely citing; "Neighborhood preserving embedding"– should be cited; Dimensionality Reduction Using the Sparse Linear Model–i didn't really get this one, but it should be cited also; Dimensionality Reduction:A Comparative Review;

Two-dimensional PCA: a new approach to appearance-based face representation and recognition Sparsity preserving projections with applications to face recognition Learning With $\ell^1$-Graph for Image Analysis

In general, I think that it's worth thinking a bit about how interested I am in classification (supervised learning) vs unsupervised/exploratory data analysis.

# 2  Applications

Possible things to look at:

1. Cardiac imaging–look at which areas of stenosis/wall thinness are most correlated with some disease outcome

2. Combining imaging and non-imaging information–Rueckert's recent paper in MedIA, looking at genomics, building prediction models for areas of brain based on genomic clusters

3. Methodology–explicit contiguity constraints. I think this may be more important than it appears . . . also maybe look at fused lasso-type penalties, but in a multivariate matrix decomposition (or for that matter, totally unsupervised context)– don't think that's been done before.

4. breast imaging (despina)? may be worthwhile . . . but not sure if it makes sense . . . maybe replace her ROI method with sparse PCA, looking for different textures and comparing against disease outcomes. [8] does with breast imaging (classification of MRI of breasts).

# Literature Cited

[1] N. Batmanghelich, B. Taskar, and C. Davatzikos. A general and unifying framework for feature construction, in image-based pattern classification. *Information processing in medical imaging : proceedings of the ... conference*, 21:423–434, 2009.

[2] N.K. Batmanghelich, B. Taskar, and C. Davatzikos. Generative-discriminative basis learning for medical imaging. *IEEE Transactions on Medical Imaging*, 31(1):51–69, 2012.

[3] O. Chapelle, P. Haffner, and V.N. Vapnik. Support vector machines for histogram-based image classification. *Neural Networks, IEEE Transactions on*, 10(5):1055 –1064, September 1999.

[4] Rmi Cuingnet, Charlotte Rosso, Marie Chupin, Stphane Lehricy, Didier Dormont, Habib Benali, Yves Samson, and Olivier Colliot. Spatial regularization of SVM for the detection of diffusion alterations associated with stroke outcome. *Medical image analysis*, 15(5):729–737, October 2011. PMID: 21752695.

[5] David Donoho and Victoria Stodden. When does Non-Negative matrix factorization give a correct decomposition into parts? In Sebastian Thrun, Lawrence Saul, and Bernhard Schlkopf, editors, *Advances in Neural Information Processing Systems 16*. MIT Press, Cambridge, MA, 2004.

[6] Xin Geng, De-Chuan Zhan, and Zhi-Hua Zhou. Supervised nonlinear dimensionality reduction for visualization and classification. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 35(6):1098 –1107, December 2005.

[7] Zhiqiang Lao, Dinggang Shen, Zhong Xue, Bilge Karacali, Susan M. Resnick, and Christos Davatzikos. Morphological classification of brains via high-dimensional shape transformations and machine learning methods. *NeuroImage*, 21(1):46–57, January 2004.

[8] J. Levman and A.L. Martel. Computer-aided diagnosis of breast cancer from magnetic resonance imaging examinations by custom radial basis function vector machine. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 5577 –5580, September 2010.

[9] B. Raducanu and F. Dornaika. A supervised non-linear dimensionality reduction approach for manifold learning. *Pattern Recognition*, 45(6):2432–2444, 2012.

[10] S.T. Roweis and L.K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[11] A. Saxena, A. Gupta, and A. Mukerjee. *Non-linear dimensionality reduction by locally linear isomaps*, volume 3316. 2004.

[12] B. Schlkopf, S. Mika, C.J.C. Burges, P. Knirsch, K.-R. Mller, G. Rtsch, and A.J. Smola. Input space versus feature space in kernel-based methods. *IEEE Transactions on Neural Networks*, 10(5):1000–1017, 1999.

[13] J.B. Tenenbaum, V. De Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

[14] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik. Dimensionality reduction: A comparative review. *Published online*, 10(February):1–35, 2007.

[15] Prashanthi Vemuri, Jeffrey L Gunter, Matthew L Senjem, Jennifer L Whitwell, Kejal Kantarci, David S Knopman, Bradley F Boeve, Ronald C Petersen, and Jr Jack, Clifford R. Alzheimer's disease diagnosis in individual subjects using structural MR images: validation studies. *NeuroImage*, 39(3):1186–1197, February 2008. PMID: 18054253.

[16] Daniela M Witten and Robert J Tibshirani. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical applications in genetics and molecular biology*, 8(1):Article28, 2009. PMID: 19572827.

[17] Robin Wolz, Paul Aljabar, Joseph V Hajnal, Jyrki Lotjonen, and Daniel Rueckert. Nonlinear dimensionality reduction combining MR imaging with non-imaging information. *Medical Image Analysis*, 16(4):819–830, May 2012. PMID: 22244037.

[18] Ming-Hsuan Yang. Extended isomap for classification. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 3, pages 615 – 618 vol.3, 2002.

[19] Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):4967, 2006.