AFRL-RI-RS-TR-2015-134

INTELLIGENT CLASSIFICATION IN HUGE HETEROGENEOUS DATA SETS

JUNE 2015

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2015-134 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/ S /
JOSEPH A. CAROLI, Chief
High Performance Systems Branch

/ S /
MARK H. LINDERMAN
Technical Advisor, Computing
& Communications Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JUNE 2015	FINAL TECHNICAL REPORT	JUL 2013 – APR 2015

4. TITLE AND SUBTITLE

INTELLIGENT CLASSIFICATION IN HUGE HETEROGENEOUS DATA SETS

5a. CONTRACT NUMBER
IN-HOUSE

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
62788F

6. AUTHOR(S)

Ashley Prater

5d. PROJECT NUMBER
93SF

5e. TASK NUMBER
IN

5f. WORK UNIT NUMBER
HO

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/RITB
525 Brooks Road
Rome NY 13441-4505

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RITB
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2015-134

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. PA# 88ABW-2015-2633
Date Cleared: 26 MAY 2015

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This research effort sought to develop basic mathematical models and algorithms to perform classification and pattern recognition on large, heterogeneous data sets. To handle the large amounts of data, the principal investigator employed sparse representations to a two-fold effect. First, the effective dimensionality of the data is greatly reduced, allowing the discovery of meaningful connections among disparate data which is essential for classification. Second, the data dimension reduction technique used lends itself well to fast processing and accurate algorithms. To this end, several research directions were pursued. An efficient algorithm for approximating Dantzig selectors, which provide sparse minimal l1-norm vectors solving a linear regression problem, is presented. The Dantzig selector model and algorithm is then extended to incorporate overcomplete dictionaries, which allows one to explore data separation, classification, and pattern recognition in heterogeneous data sets. Finally, new research is presented approximating high dimensional generalized Fourier series with a focus upon using the series for a novel method to perform rotational invariant pattern recognition in images.

15. SUBJECT TERMS
Machine Learning, Supervised Classification, Principal Component Analysis, Compressed Sensing, Pattern Recognition

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**ASHLEY PRATER**
U	U	U	UU	12	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

Contents

1 Summary

The purpose of this in-house research project was to investigate supervised and semi-supervised machine learning techniques for classification and pattern recognition by exploiting the natural sparsity in signals and through data dimension reduction, and to develop and tailor algorithms for the extraction of intelligence from several huge heterogeneous data sets. The research provides a mathematically rigorous foundation for models and algorithms that could be applied toward technologies in the areas of autonomy, trusted systems, situational awareness and machine guided data-to-decision processes.

This report describes the delivered results for three related research areas. First, an algorithm to approximate the Dantzig selector is described. The Dantzig selector is a method to approximate a sparse vector that captures the most essential information in a large amount of data. The algorithm, which was developed in the course of this research effort, is an iterative approach based upon solutions to a pair of proximity operator equations. The algorithm is an improvement over current state-of-the-art methods in that it produces results of similar quality, but tends to converge significantly faster. Next, an ℓ_1 minimization model is extended to incorporate overcomplete dictionaries. The extension allows one to obtain a sparse representation of homogeneous and heterogeneous data, which in turn is used to improve classification and pattern recognition using the sparse coefficient vectors. Additionally the proposed method is demonstrated to separate composite signals using a supervised machine learning technique. Finally, an unique method to perform rotational invariant pattern recognition is described. The method is based upon an efficient strategy for approximating the Gaussian-Hermite moments of a function using a collocation-based optimization approach. The method described herein is an improvement in accuracy and complexity over the commonly used brute-force approaches.

2 Introduction

Machine learning is the field concerning the conversion of data into usable information by a computer through the discovery of patterns and trends that are present in the data, but are typically difficult for a human to discern due to the sheer mass of the data and nuanced interactions between variables in the feature space. A general machine learning model seeks to label the high dimensional input data with the appropriate output classifiers. That is, if $\mathbb{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots | \mathbf{x}_i \in \mathbb{R}^d\}$ is a collection of d-dimensional real-valued input vectors and $\mathbb{Y} = \{y_1, y_2, \dots, y_k | y_i \in \mathbb{R}\}$ is a collection of real-valued output labels, a machine learning process attempts to find a well-defined function $f : \mathbb{X} \to \mathbb{Y}$ that accurately matches each $\mathbf{x}_i$ with its appropriate label $f(\mathbf{x}_i) = y_j$. Machine learning techniques can follow the supervised, unsupervised or semi-supervised paradigms.

Supervised machine learning techniques are used when the user has some known ground truth pairs $(\mathbf{x}, y)$ available. The set of vectors with known classifiers is called the training set, and this knowledge is exploited to intelligently extrapolate function pairs $f(\mathbf{x}) = y$ for vectors $\mathbf{x}$ that are not encountered in the training set. Some specific approaches that are categorized as supervised learning techniques include support vector machines, decision trees, and the training of artificial neural networks, where each approach admits a number

of specific algorithms. A standard example of supervised learning techniques is OCR (optical character recognition), which was popularized by the United States Postal Service, but supervised learning techniques have applications in any general pattern recognition environment where a known training set is available. A properly sampled training set can prepare the system for success, however reliance on a training set that does not accurately capture the array of data one might encounter can lead to misclassification and unreliable results. Additionally, the algorithms used in supervised learning work best for data that can be expressed in a features space of relative small effective dimension with linear relations among the features.

On the other hand, unsupervised machine learning techniques classify the input data according to similar structures that reveal themselves through mathematical data dimension reduction and feature extraction. These techniques are employed when no training set is available, and therefore also no target output attributes are known. Unsupervised learning algorithms cluster the data into sets that exhibit similar properties. Common approaches to unsupervised machine learning include k-means clustering algorithms, and can be achieved through a variety of methods including PCA (principal component analysis) and SVD (singular value decomposition) techniques. The absence of a training set and target classifier can offer advantages over supervised learning; allowing the data to reveal their own correlations instead of having user-imposed restrictions on classification labels can provide richer intelligence from the information.

The following described research draws from supervised and unsupervised machine learning paradigms, with an emphasis on reducing the effective dimension of the data by finding accurate sparse representations of the data.

3 Methods, Assumptions, and Procedures

The objective of this research is to investigate machine learning techniques for classification and pattern recognition, and to develop and tailor adaptive algorithms for application to huge, heterogeneous data sets enabling the extraction of intelligence from information.

The approach uses two-scale supervised and unsupervised machine learning techniques to discover inter- and intra-set relationships and to reduce the dimensionality of the data.

Unsupervised data dimension reduction through the use of ℓ_1 norm minimization and PCA will be performed within each data set and across multiple sets to harvest the most signification features while suppressing spurious information. The techniques used ensures minimal redundancy of the representation of the data, and allows one to discover significant interactions among the relevant features in each set and to enable better situational awareness.

For the supervised learning aspect of the research, we will begin with a large collection of known information, and analyze the characterizing relationship among the data in each class. This will lead to a sufficiently large set of ground truth, as well as giving guidance on the number of classes to be used, and the correct classification function f. These cannot be known a priori - access to initial information is fundamental in developing a supervised learning scheme. However, if the relationships in the data tend to be described well by a linear classifier, the PCA technique can be used again in the discovery of f. For example,

if the principle component of x is p, and x is known to belong to class y, then any newly encountered data with principle component p shall also be assigned to class y.

The algorithms will be developed and implemented using MATLAB. Speed and memory usage may be improved by converting the algorithms to C.

4 Results and Discussion

Several new methods and algorithms were developed during the course of this research effort, including an algorithm to quickly and accurately approximate the Dantzig selector [6], a scheme to separate and classify undersampled composite data [7], and a novel method to perform rotational invariant pattern recognition in images [8]. These three major research products have been summarized in journal publications and conference proceedings. To illustrate and verify the theoretical results of the main research products listed above, several interesting numerical experiments were performed using real-world and simulated data with large, homogeneous and heterogeneous data sets. Each is explained in more detail below.

4.1 Algorithm to Approximate the Dantzig Selector

The Dantzig selector is a solution to the optimization problem

$$\hat{\beta} \in \underset{\beta}{\operatorname{argmin}} \left\{ \|\beta\|_1 : \|D^{-1}X^\top(X\beta - y)\|_\infty \leq \delta \right\}, \tag{1}$$

where y is the observed data, X is a known data matrix satisfying certain properties [4], D is a diagonal matrix normalizing the columns of X, and δ is a small, user-chosen parameter. Typically the dimensionality of β is much larger than that of y, however since the solution to Equation (1) tends to be sparse, the Dantzig selector has a much lower effective dimensionality that the original data. Therefore the Dantzig selector is an appropriate feature to use for data dimension reduction to assist machine learning and classification tasks.

Several methods exist to compute a Dantzig selector, including a primal-dual interior point method [3, 4], a first-order method based upon linear cone programming [1, 2], and an alternating direction method [5]. Each has its own strength and weaknesses. For example, the Alternating Direction Method of Multipliers is an iterative approach that tends to converge to a solution of (1) in few iterations, however the total computational cost of this method is large since each step in the iteration requires the solution of another optimization problem via an iterative approach. To solve the Dantzig selector problem, Lixin Shen (SU), Bruce Suter (AFRL/RITB) and Ashley Prater (AFRL/RITB) developed a two-stage approach. The first stage approximates $\hat{\beta}$ as the fixed point solution to a pair of iterative proximal equations. This fixed point solution tends to very accurately recover the support of the Dantzig selector, but allows errors in the magnitude of the nonzero entries. The second stage, a postprocessing step, corrects this issue by regressing the observed data onto the support of the fixed point solution.

In further detail, the optimization problem (1) can be expressed as

$$\hat{\beta} \in \underset{\beta}{\operatorname{argmin}} \left\{ \|\beta\|_1 + \iota_{\mathcal{C}}(A\beta) \right\}, \tag{2}$$

where $A = D^{-1}X^\top X,\ b = D^{-1}X^\top y,\ \mathcal{C} = \{\beta : \|\beta - b\|_\infty \leq \delta\}$ and $\iota_{\mathcal{C}}(\cdot)$ is an indicator function on the set $\mathcal{C}$. Solutions to Equation (2) can be characterized by β and τ satisfying

$$\begin{cases} \beta &= \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}\left(\beta - \frac{\lambda}{\alpha}A^\top\tau\right), \\ \tau &= \left(I - \operatorname{prox}_{\iota_{\mathcal{C}}}\right)(A\beta + \tau), \end{cases} \tag{3}$$

where, for a function f with parameter λ, the proximity operator is defined by

$$\operatorname{prox}_{\lambda f}(x) := \underset{u\in\mathbb{R}^d}{\operatorname{argmin}} \left\{ \frac{1}{2\lambda}\|u - x\|_2^2 + f(u) \right\}.$$

The iterative method developed as part of this research effort approximates a solution of (3) as the fixed point solution of

$$\begin{cases} \tau^{k+1} &= \operatorname{prox}_{\delta\|\cdot\|_1}\left(A(2\beta^k - \beta^{k-1}) + \tau^k - b\right), \\ \beta^{k+1} &= \operatorname{prox}_{\frac{1}{\alpha}\|\cdot\|_1}\left(\beta^k - \frac{\lambda}{\alpha}A^\top\tau^{k+1}\right), \end{cases} \tag{4}$$

for $k = 1, 2, 3\ldots$. The fixed point solution of (4) can be solved straightforwardly using a soft thresholding operator. The overall complexity of this iterative approach is $\mathcal{O}(np)$, where n is the dimension of the observation y and p is the dimension of the Dantzig selector $\hat{\beta}$.

The above method was compared to the popular Alternating Direction Method (ADM) for finding the Dantzig selector. We found that while the accuracy of the approximated solutions was nearly equal for the two approaches, our method was significantly faster. Notably, the difference in CPU runtime for the two approaches grew larger as the dimensionality increased and also as original data was corrupted by more noise.

To demonstrate the strength of the proposed method, Drs. Prater, Shen and Suter used a large, heterogeneous data set of biomarker data and employed the Dantzig selector as a classifier to predict whether a patient may have a future leukemia diagnosis. The dataset included numerical data for over 7000 biomarkers. It is likely that only a small number of genes will contribute to the likelihood of a patient developing leukemia in the future, but it is a difficult problem for even a medically trained individual to identify the contributing genes from the huge amount of data. We found that the Dantzig selector performed well in determining the the small number of biomarkers from this dataset that contribute most to a patient developing leukemia. The Dantzig selector was used in the following manner. To train the output classes, we used a simple supervised machine learning approach. A portion of the dataset was designated as ground truth, and the corresponding values of the observed data vector y were set equal to 1 if a patient had a leukemia diagnosis and 0 otherwise. The trained value of the Dantzig selector $\hat{\beta}$ was then used with the unclassified test data to compute the output observation y_{test}. Finally, the diagnosis of the patients in the test category were predicted using a clustering method with two classes.

The method proposed above took on average less than 0.01 seconds for most parameter selections and yielded only one misdiagnosed patient out of 204 trials. In comparison, the ADM required more than 100 seconds on average and misdiagnosed 7 patients out of the 204 trials.

Futher details on the discussion above can be found in [6].

4.2 Scheme to Separate and Classify Composite Data

After developing the algorithm described above to approximate the Dantzig selector, attention was turned to extending both the model and the algorithm to accept more general types of signals. The Dantzig selector model in Equation (1) yields a sparse solution $\hat{\beta}$ only for observations y admitting such a representation. This is a rather restrictive class of signals. Instead, we looked to incorporate into the model not only a representation basis, but overcomplete dictionaries so one could use the model to analyze a rich class of images and signals.

To this end, suppose that c is the signal or image one wishes to analyze, and that it is comprised of several other atomic signals. Say, $c = c_1 + c_2 + \cdots + c_k$, where each individual component c_j is unknown. In applications it is unlikely that the individual components are sparse, but each one could admit a sparse representation $c_j = B_j\beta_j$, where β_j is a sparse vector and B_j is a basis or dictionary. The B_js could possibly coincide. In practice one does not know a priori the sparse vectors β_j, but typically one knows a good sparsifying basis B_j. Given the observation $y = Xc$, where X is defined as in (1), one can incorporate these bases into the model as

$$\hat{\beta} = \operatorname{argmin} \left\{ \|\beta\|_1 : \left\| D^{-1}B^\top X^\top \left(XB\beta - y \right) \right\|_\infty \leq \delta \right\}, \tag{5}$$

with the matrix B equal to the concatenation of the bases and the vector β equal to the concatenations of the sparse vectors β_j.

The proximity operator fixed point based algorithm described above can be easily extended to include these overcomplete dictionaries by redefining A, b and $\mathcal{C}$ as

$$A = D^{-1}B^\top X^\top XB, \quad b = D^{-1}B^\top X^\top y, \quad \text{and} \quad \mathcal{C} = \{\beta : \|b - \beta\|_\infty \leq \delta\}.$$

In [7], several numerical experiments illustrate the effectiveness of the scheme incorporating the overcomplete dictionaries into the Dantzig selector model. One experiment in particular is interesting in that it is paired with supervised machine learning techniques to perform separation and classification of images using the principal components of the Dantzig selectors of the components. In the example compositions of two handwritten digits arseparated and classified. The handwritten digits are taken from the United States Postal Service data set [10], which was then split into a training set and a testing set. The overcomplete dictionaries were then formed using the principal components from the labeled examples from each class included in the training set. That is, each B_j was formed by the first k principal components of the collection R_j, where R_j was all training images in the j^{th} class. Supposing that c was a composition of two unlabeled digits taken from the testing dataset, the Dantzig selector of c was computed from Equation (5). The unknown components of c can then be immediately recovered and classified through the Dantzig selector, based on the the support of $\hat{\beta}$. An illustration of this method is shown in Figure 1.

Further details on the discussion above can be found in [7].

4.3 Method to Perform Rotational Invariant Pattern Recognition

Under this research effort, a mathematically rigorous method to perform pattern recognition in noisy, possibly rotated images was developed that computes feature vectors using a sparse

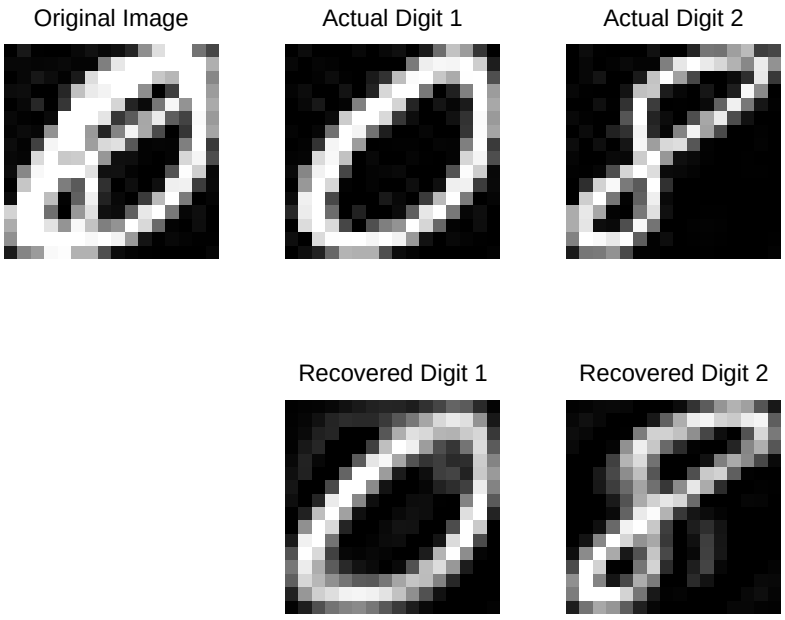

Figure 1: A composition of unknown handwritten digits separated using the Dantzig selector with trained overcomplete dictionaries.

representation of the data. The approach was to write the unclassified two dimensional image in terms of the bivariate Hermite polynomials, say

$$f(x,y) \approx \sum_{(n_1,n_2)\in W} c_{(n_1,n_2)} H_{(n_1,n_2)}(x,y), \tag{6}$$

where f describes the image, $H_{(n_1,n_2)}$ is the $(n_1,n_2)-$th Hermite polynomial, $c_{(n_1.n_2)}$ are real valued coefficients, and W is an appropriately chosen index set. The Hermite polynomials are used in the Fourier-like expansion (6) because certain combinations of their moments are rotation invariant. The (n_1,n_2) geometric Gaussian-Hermite moment of the image f is defined by

$$m_{(n_1,n_2)} := \iint_{\mathbb{R}^2} f(x,y) H_{(n_1,n_2)}(x,y) e^{-(x^2+y^2)/2} \, \mathrm{d}x \, \mathrm{d}y. \tag{7}$$

In [8], it is shown that for carefully chosen W, the $(n_1,n_2)-$th geometric Gaussian-Hermite moment of f can be well approximated by the $(n_1,n_2)-$th coefficient appearing in the expansion (6).

Computing the coefficients appearing in (6) is nontrivial. Each one is defined by a highly oscillatory bivariate integral for which closed-form solutions exist in only rare cases. Each is difficult to compute either directly or using quadrature methods, and one must perform this approximation as many times as the cardinality of the index set W. To quickly and accurately approximate the coefficients, and therefore also the geometric Gaussian-Hermite moments of an image, Dr. Prater proposed in [8] using a sparse collocation-based approach.

Suppose X is the Jacobi-like matrix defined by $X(j,k) = H_k(x_j)$, with the univariate Hermite polynomial H_k and predetermined nodes $\Lambda = \{x_1, x_2, \ldots, x_m\}$. Then one can approximate the entire collection of coefficients $\{c_{(n_1,n_2)}\}$ appearing in (6) by solving the optimization problem

$$\begin{cases} \text{minimize} & \|c\|_1 \\ \text{subject to} & \|D^{-1}X^\top(Xc - f)\|_\infty \leq \delta. \end{cases} \tag{8}$$

Suppose $\hat{c}$ is the solution to (8). One can approximate the rotation invariants of the Gaussian-Hermite moments from $\hat{c}$ using the equations documented in [8, 9]. The first few rotation invariants of the Gaussian-Hermite moments are given by:

$$\phi_1 = \hat{c}_{(2,0)} + \hat{c}_{(0,2)},$$
$$\phi_2 = \left(\hat{c}_{(3,0)} + \hat{c}_{(1,2)}\right)^2 + \left(\hat{c}_{(0,3)} + \hat{c}_{(2,1)}\right)^2,$$
$$\phi_3 = \left(\hat{c}_{(2,0)} - \hat{c}_{(0,2)}\right)\left[\left(\hat{c}_{(3,0)} + \hat{c}_{(1,2)}\right)^2 - \left(\hat{c}_{(0,3)} + \hat{c}_{(2,1)}\right)^2\right] + 4\hat{c}_{(1,1)}\left(\hat{c}_{(3,0)} + \hat{c}_{(1,2)}\right)\left(\hat{c}_{(0,3)} + \hat{c}_{(2,1)}\right).$$

To perform rotational invariant pattern recognition, we clasify images according to how closely the collection of the rotation invariants match those of labeled test images. That is, suppose $\{\Phi_1, \Phi_2, \ldots, \}$ are vectors of rotation invariants of the Gaussian-Hermite moments of the labeled images $\{F_1, F_2, \ldots\}$, and let Φ be the rotation invariants of the Gaussian-Hermite moments of the unclassified image f computed using method (8). Then classify the image f as a rotation of image F_j if

$$\|\Phi_j - \Phi\|_1 \leq \|\Phi_k - \Phi\|_1, \ \forall k. \tag{9}$$

The method described above is computationally superior to more direct 'brute-force' style methods. The direct method would have several rotations of each example images included in the training data set, resulting in more differences and comparisons to make in Equation (9).

For more details on the above, including numerical experiments using real-world and simulated noise-free and noisy data, see [8]. The above work demonstrated this research can be an effective pre-processing step and classification strategy for certain machine learning tasks. This research sub-project will continue to be studied in the Neuromorphic Computing group at AFRL/RI.

5 Conclusions

Throughout the research project, models and algorithms were explored and developed that can be used to perform the supervised classification and pattern recognition in large, heterogeneous data sets. The research was broad in scope and has direction applicability in several data domains, including the previously stated area of recognition vehicles from several data sensor products. The PI will continue to pursue these directions in new research efforts.

6 References

[1] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM J. Imaging Sci. 2 (2009), 183–202.

[2] S. Becker, E. Candes and M. Grant, *Templates for convex cone problems with applications to sparse signal recovery*, Math. Program. Comput. 3, 165–218.

[3] S. Boyd and L. Vandenberghe, *Convex Optimization.* Cambridge University Press. 2004.

[4] E. Candes and T. Tao, *The Dantzig selector: Statistical estimation when p is much larger than n*, The Annals of Statistics, Volume 35, Number 6 (2007) pp. 2313–2351.

[5] Z. Lu, T.P. Pong and Y. Zhang, *An alternating direction method for finding Dantzig selectors,* Comput. Statist. Data Anal. 56 (2012), 4037–4046.

[6] A. Prater, L. Shen and B. W. Suter, *Finding Dantzig selectors with a proximity operator based fixed-point algorithm*, Computational Statistics and Data Analysis, Available online 23 April 2015. DOI: 10.1016/j.csda.2015.04.005

[7] A. Prater and L. Shen, *Separation of undersampled composite signals using the Dantzig selector with overcomplete dictionaries*, IET Signal Processing, Available online 17 April 2015. DOI: 10.1049/iet-spr.2014.0048

[8] A. Prater, *Sparse generalized Fourier series via collocation-based optimization*, 2014 IEEE Applied Imagery and Pattern Recognition Workshop Conference Proceedings, pp. 1–8. DOI: 10.1109/AIPR.2014.7041926

[9] B. Yang and M. Dai, *Image analysis by Gaussian-Hermite moments*, Signal Processing, 91 (2011) 2290–2303.

[10] *Data for MATLAB hackers*, http://www.cs.nyu.edu/roweis/data.html. Accessed 05/16/14.

A List of Acronyms

ADM	Alternating Direction Method
AFRL	Air Force Research Laboritory
CCD	Coherent Change Detection
CPU	Central Processing Unit
CTC	Core Technical Competencies
DoD	Department of Defense
GMTI	Ground Moving Target Indicator
ISR	Intelligence, Surveillance and Reconnaissance
NCD	Noncoherent Change Detection
OCR	Optical Character Recognition
PCA	Principal Component Analysis
SAR	Synthetic Aperture Radar
SVD	Singular Value Decomponsition
USPS	United States Postal Service