# AM221 Final Project Proposal

Taylor Killian & Leonhard Spiegelberg

April 29, 2016

**Abstract**

To be filled with an interesting summary of our work and results

## 1 Dictionary Learning as 2-stage supermodular minimzation

The introduced problem of Dictionary learning, which in its general form can be seen as

$$\min_{D,R,\theta,\lambda} f(D,R,X) + g(D,\theta) + h(R,\lambda)$$

where $f$ describes the objective function used to measure goodness of approximation of $X$ through $D, R$, $g$ describing suitable constraints on the dictionary, $h$ on the representation respectively. With an input dataset $X = [x_1, \ldots, x_k], x_i \in \mathbb{R}^d, X \in \mathbb{R}^{d\times k}$ we wish to find a dictionary $D \in \mathbb{R}^{d\times n}, \mathcal{D} = [d_1, \ldots, d_n]$ and a representation $\mathcal{R} = [r_1, \ldots, r_k],\ r_i \in \mathbb{R}^n,\ \mathcal{R} \in \mathbb{R}^{n\times k}$, such that both $\|X - \mathcal{D}\mathcal{R}\|_F^2$ is minimized and the representations $r_i$ are "sparse enough". To limit the dictionary becoming infinitely large (or small) we introduce a constraint on the dictionary's columns. For this any sufficient norm can be used. An often used norm is the $l_2$-norm. Thus for the dictionary learning problem we introduce the problem with constraints

$$\min_{D,R} \|X \ - DR\|_F^2 \tag{1}$$

$$\text{s.t.} \qquad \|d_j\|_2 \leq 1, \forall j = 1, ..., n \tag{2}$$

$$\|r_i\|_0 \leq t, \forall i = 1, ..., k \tag{3}$$

Description of constraints here

Using lagrange multipliers this can be brought to the general form above

$$\min_{D,R,\theta,\lambda} \|X \ - DR\|_F^2 + \sum_{j=1}^{m} \theta_j(\|d_j\|_2 - 1) + \sum_{i=1}^{k} \lambda_i(\|r_i\|_0 - t)$$

I.e. the functions are

$$
\begin{aligned}
f(D, R, X) &:= \|X - DR\|_F^2 \\
g(D, \theta) &:= \sum_{j=1}^{m} \theta_j(\|d_j\|_2 - 1) \\
h(R, \lambda) &:= \sum_{i=1}^{k} \lambda_i(\|r_i\|_0 - t)
\end{aligned}
$$

How to bring this to the penalty function form?

# 2 Introduction

Introduce the problem and it's complexities

# 3 Background Information

## 3.1 Background

Here we'll want to add some information about Sparse regression, set functions and definitions of sub modular and super modular, Dictionary Selection, Greedy Methods, etc.

## 3.2 Related Work

A simple overview of prior work

# 4 Methods

Place holder section to begin to outline our work

**Algorithm 1** Notional algorithm

```
Figure out optimal arrangement of rows and columns of input data
while there's still time in the semester do
    if f(p_k + s_k λ_i) < f(p_k) for some λ_i then
        p_{k+1} = p_k + s_k λ_i
        s_{k+1} = s_k
    else
        p_{k+1} = p_k
        s_{k+1} = α s_k
    end if
end while
return VICTORIOUS
```

# 5 Experiments

Hold for applying our methods

## 5.1 Results

The results of our algorithms against others

# 6 Conclusions & Future Work

## 6.1 Conclusions

We will have done it!

## 6.2 Future Work

Rule the world

# References


[1] Das, A., Kempe, D., (2011). Submodular meets Spectral: Greedy Algorithms for Subset Selection, Sparse Approximation and Dictionary Selection. *Proceedings of the 28th International Conference on Machine Learnings.*

[2] Cevher, V., Krause, A., (2011). Greedy Dictionary Selection for Sparse Representation. *Selected Topics in Signal Processing, IEEE Journal of 5* (5), pp. 979 - 988.

[3] Doshi-Velez, F., Williamson, S., (2015). Restricted Indian Buffet Processes. *In submission.* arXiv:1508.06303 [stat.ME]