
Sparse Coding for Dictionary Learning in Context of Image De-noising

Dhaivat Deepak Shah
ds3267@columbia.edu

Gaurav Ahuja
ga2371@columbia.edu

Sarah Panda
sp3206@columbia.edu

Abstract

Dictionary learning involves solving the following optimization problem: $\min_x \|x\|_1 \text{ s.t. } y = Dx$ where y is the input signal, D is the dictionary and x is the sparse representation of the signal. The problem of image restoration has been addressed with a multitude of approaches. All the approaches to solve the optimisation problem fall under the 3 broad categories of Relaxation (Basis Pursuit), Greedy approach (Matching Pursuit) or Hybrid methods. Our project primarily focuses on the Relaxation methodology. Here, both y and D are unknown. Mairal, Julien, et al, 2009 present an online learning algorithm [1] which involves two optimization problems. First, assumes y to be available and D minimizes over D . This is known as the sparse coding problem. Second, updates y after D obtaining y . Mairal, Julien, et al, 2009 use LARS [2] to solve the sparse coding problem. We propose to compare the performance of the online dictionary learning algorithm by solving the sparse coding problem using methods [1][5] like feature sign [3], FISTA [4], Interior point, Sequential Shrinkage or Iterative Shrinkage methods and Stochastic Gradient Descent in the context of image restoration.

1 Introduction

Introduction : Problem of Image denoising

2 Intro to Dictionary learning

Dictionary Learning is a focus area in the Signal Processing domain. A dictionary is used for the sparse representation or approximation of signals. A dictionary D , is an $N \times K$ matrix that contains K prototype signals called atoms. A signal y_0 can be represented by a linear combination of a small number of these atoms. For any signal y_0 and dictionary D there must exist a sparse vector $x \in R^K$ such that:

$$y_0 = D * x \quad (1)$$

There will be cases where we cannot exactly represent the signal with a linear combinations of atoms from a finite dictionary. In such a case we find an approximation of the signal, y such that:

$$y = D * x \quad (2)$$

$$\|y - y_0\| < \epsilon \quad (3)$$

Dictionary Learning is a problem of finding a dictionary such that the approximation of many vectors in the training set are as good as possible.

For a finite number of training vectors, L each of length N the aim of Dictionary Learning is to find both a Dictionary D , of size $N \times K$ and the L corresponding coefficients vectors of length K .

In this paper we will apply the problem of Dictionary Learning on noisy image. The image is broken into 8×8 pixel patches and this $N = 64$ pixel vector is approximated by the linear combination of atoms of the Learned Dictionary.

2.1 Problem Statement

Classical dictionary learning techniques consider a finite training set of signals $Y = [y_1, y_2, \dots, y_L]$ in $\mathbb{R}^{N \times L}$ and optimize the empirical cost function:

$$f(D) = \frac{1}{L} \sum_{i=1}^L l(y_i, D) \quad (4)$$

where $D \in \mathbb{R}^{N \times K}$ is the dictionary and l is the loss function such that $l(y, D)$ is small if D is a good at representing the signal x .

$l(y, D)$ can be define as the optimal value of the l_1 or l_0 sparse coding problem. See section on Sparse coding for more details, TODO: rephrase this. We consider the l_1 norm for illustration in this section.

$$l(y, D) = \min_{x \in \mathbb{R}^K} \frac{1}{2} \|y - Dx\|_2^2 + \lambda \|x\|_1 \quad (5)$$

λ is the regularization parameter. The problem of minimizing the empirical cost function $f(D)$ is not convex with respect to D . It can be rewritten as a joint optimization problem with respect to dictionary D and the coefficients $x = [x_1, x_2, \dots, x_L]$, which is not jointly convex, but convex with respect to each of the two variables D and x when the other one is fixed.

$$\min_{D \in \mathbb{R}^{N \times K}, x \in \mathbb{R}^{K \times L}} \frac{1}{L} \sum_{i=1}^L \left(\frac{1}{2} \|y_i - Dx_i\|_2^2 + \lambda \|x_i\|_1 \right) \quad (6)$$

A natural approach to solving this problem is to alternate between the two variables, minimizing one over the other while keeping the other fixed.

3 KSVD

KSVD for learning dictionaries

4 Sparse Coding

Sparse coding problem explained in deep and ways to approximate the sparse code - Basis pursuit - Matching pursuit

5 Summary of sparse coding techniques used:

5.1 FISTA

- FISTA

5.2 MP

- MP

5.3 OMP

- OMP

5.4 ALM

- ALM

5.5 Feature Sign

Feature sign aims to solve the following L-1 regularized sparse coding problem

$$\min_x f(x) = \|y - Dx\|^2 + \lambda \|x\|_1 \quad (7)$$

The intuition behind this algorithm is that if we know the signs(positive, negative or zero) of x_i at the optimal value we can replace $\|x_i\|$ by either x_i if $x_i > 0$, $-x_i$ if $x_i < 0$ or 0 if $x_i = 0$. Considering only non zero coefficients this problem reduces to a standard quadratic optimization problem. This problem can then be solved analytically and efficiently. Feature sign algorithm tries to search for the signs of the coefficients. Given any such search results, it efficiently solves the resulting unconstrained quadratic problem. Further the algorithm systematically refines the guess if it turns out to be initially incorrect.

1. **Initialize** $x := 0$, $\theta : 0$ and active set $:= \{\}$, $\theta_i \in \{-1, 0, 1\}$ denotes $\text{sign}(x_i)$
2. From zero coefficients of x , select $i = \text{argmax}_i \left| \frac{\partial \|y - Dx\|^2}{\partial x_i} \right|$
 Activate x_i (add i to active set) only if it locally improves the objective
 - (a) If $\left| \frac{\partial \|y - Dx\|^2}{\partial x_i} \right| > \lambda$; set $\theta_i := -1$, active set $:= \{i\} \cup \text{active set}$
 - (b) If $\left| \frac{\partial \|y - Dx\|^2}{\partial x_i} \right| < -\lambda$; set $\theta_i := 1$, active set $:= \{i\} \cup \text{active set}$
3. **Feature-Sign step:**
 Let \hat{D} be a submatrix of D that contains only the columns corresponding to the active set.
 Let \hat{x} and $\hat{\theta}$ be sub vectors of x and θ corresponding to the active set.
 Compute the analytical solution to the resulting quadratic problem
 $\hat{x}_{new} := (\hat{D}^T \hat{D})^{-1} (\hat{D}^T y - \lambda \hat{\theta} / 2)$
 Perform a discrete line search on the closed line segment from \hat{x} to \hat{x}_{new} :
 - Check the objective value at \hat{x}_{new} and all points where any coefficient changes sign.
 - Update \hat{x} (and all corresponding entries in x) to the point with the lowest objective value
 Remove zero coefficients of \hat{x} from the active set and update $\theta := \text{sign}(x)$
4. **Check the optimality conditions:**
 - (a) Optimality condition for nonzero coefficients: $\left| \frac{\partial \|y - Dx\|^2}{\partial x_i} \right| + \lambda \text{sign}(x_i) = 0 \forall x_i \neq 0$
 If this condition is not satisfied, go to Step 3(without any new activation); else check condition (b).
 - (b) Optimality condition for zero coefficients: $\left| \frac{\partial \|y - Dx\|^2}{\partial x_i} \right| + \lambda \text{sign}(x_i) \leq 0 \forall x_i = 0$
 If this condition is not satisfied go to Step 3; otherwise return x as the solution.

5.6 TNIPM

In [2], the authors formulate a Truncated Newton Interior Point method. The algorithm solves the L-1 regularised problem:

$$\min_x f(x) = \|y - Dx\|^2 + \lambda \|x\|_1 \quad (8)$$

This problem can be transformed to a convex quadratic program with linear inequality constraints,

$$\min \|y - Dx\|^2 + \sum_{i=1}^K \lambda u_i \quad (9)$$

$$\text{subject to: } -u_i \leq x_i \leq u_i, \forall i \in \{1, K\} \quad (10)$$

The aim of this algorithm is to solve this quadratic program using the Newton's method. Since solving the Newton system exactly is computationally expensive for large L-1 problems, TNIPM uses an iterative method, Preconditioned Conjugate Gradient (PCG) to approximately solve the Newton system and develops an interior point method to solve the L-1 problem.

6 Experimental Setup

7 Findings

8 Analysis

9 Conclusion

References

- [1] Michal Aharon, Michael Elad, and Alfred Bruckstein. -svd: An algorithm for designing over-complete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, 2006.
- [2] Seung-Jean Kim, Kwangmoo Koh, Michael Lustig, and Stephen Boyd. An efficient method for compressed sensing. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, volume 3, pages III–117. IEEE, 2007.
- [3] Honglak Lee, Alexis Battle, Rajat Raina, and Andrew Y Ng. Efficient sparse coding algorithms. *Advances in neural information processing systems*, 19:801, 2007.
- [4] Stéphane G Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993.
- [5] Yagyensh Chandra Pati, Ramin Rezaifar, and PS Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*, pages 40–44. IEEE, 1993.
- [6] Allen Y Yang, Zihan Zhou, Arvind Ganesh, S Shankar Sastry, and Yi Ma. Fast l1-minimization algorithms for robust face recognition. *arXiv preprint arXiv:1007.3753*, 2010.

10 Appendix

10.1 Appendix-1

10.2 Appendix2