ADMM Pursuit for Manifold Regularized Sparse Coding

Yael Yankelevsky and Michael Elad
Computer Science Department
Technion - Israel Institute of Technology
Email: {yaelyan,elad}@cs.technion.ac.il

***Abstract*—In this paper, we propose an efficient ADMM-based algorithm for graph regularized sparse coding that explicitly takes into account the local manifold structure of the data. Specifically, the graph Laplacian representing the manifold structure is used as a regularizer, encouraging the resulting sparse codes to vary smoothly along the geodesics of the data manifold. By preserving locality, the obtained representations have more discriminating power compared with traditional sparse coding algorithms and thus can better facilitate machine learning tasks such as classification and clustering. The experimental results demonstrate the effectiveness of our proposed algorithm over other previously suggested approaches, in terms of both lower representation errors and faster, more stable runtimes.**

I. Introduction

Sparse coding has become a popular paradigm for data representation and has been effective in practical computer vision tasks such as image denoising [1], inpainting [2], classification [3] and object recognition [4]. Due to the redundancy of the representative dictionary, small variations in the data may result in very distinct representations. To overcome this limitation, several sparse coding methods have been proposed that enforce structural sparsity patterns, for example by adding spatial consistency constraints [5] or a group-sparsity regularizer [6]. Another approach, motivated by the recent progress in spectral graph theory and manifold learning, is graph regularized sparse coding, which explicitly exploits the local geometrical structure of the data. The underlying assumption is that in many real applications, the data is likely to reside on or near a low-dimensional manifold embedded in the high-dimensional ambient space. Encoding the manifold structure by a graph, its Laplacian matrix L can be incorporated into the sparse coding framework as a regularizer. The added regularization limits the degree of freedom in the sparse coding task and favors solutions preserving the local geometry, i.e. varying smoothly along the geodesics of the data manifold.

II. Manifold regularized sparse coding

The graph regularized sparse coding problem is formulated as:

$$\arg\min_{X} \ \|Y-DX\|_F^2 + \gamma Tr(XLX^T) + \beta \sum_i \|x_i\|_1 \qquad (1)$$

where Y is the data matrix, X is the corresponding sparse representations matrix, D is an overcomplete dictionary with normalized columns (atoms) and L is the Laplacian matrix of the data manifold.

Due to the imposed graph constraint, the problem is no longer separable, and the sparse representations of different signals are now dependent on each other, demanding joint sparse coding of the ensemble signals (i.e. obtaining all columns of X together).

Several works have recently studied this problem. Zheng et al. [7] proposed to solve Equation (1) using a coordinate descent approach and subgradient methods. Other previously proposed methods are based on the feature sign search algorithm [8] or a modified sequential quadratic programming [9].

We propose a different solution based on the Alternating Direction Method of Multipliers (ADMM) [10], which enables simultaneous update of all columns of X. In this approach, the sparsity constraint is separated from the rest and Equation (1) is reformulated as

$$\begin{aligned} &\arg\min_{X} \ \|Y-DX\|_F^2 + \gamma Tr(XLX^T) + \beta \sum_i \|z_i\|_1 \\ &\text{s.t.} \quad X = Z. \end{aligned} \qquad (2)$$

The augmented Lagrangian is then given by

$$\mathcal{L}_\rho(X,Z,U) = f(X) + g(Z) + \rho\|X-Z+U\|_2^2, \qquad (3)$$

where $f(X) = \|Y-DX\|_F^2 + \gamma Tr(XLX^T)$, $g(Z) = \beta\sum_i \|z_i\|_1$, and U is the scaled dual form variable. The ADMM iterative solution consists of sequential optimizations of $\mathcal{L}_\rho$ over each of the variables X,Z, and U. The sub-problem of updating X is now quadratic, and by derivation reduces to a Sylvester equation [11]:

$$(D^TD+\rho I)X + \gamma XL = D^TY + \rho(Z-U). \qquad (4)$$

Since the eigenvalues of $(D^TD+\rho I)$ and $(-\gamma L)$ are distinct, a unique solution X is guaranteed [12]. A numerical solution can be efficiently obtained using the Bartels-Stewart algorithm [13], [14], based on a Schur decomposition and backward substitution. Alternatively, for large dimensions, an iterative gradient descent approach may be applied.

The sub-problem of updating Z reduces to a shrinkage problem, requiring applying soft thresholding to $X+U$. We denote this operator by $\mathcal{P}_{\frac{\beta}{2\rho}}$.

The graph regularized sparse coding algorithm is summarized in Algorithm 1. To speed the convergence, X is initialized with the standard sparse coding. We note that in [15] we have proposed a similar algorithm using an ℓ_0 sparsity constraint. For the ℓ_0 setting, the soft-thresholding here performed for updating Z is replaced with a hard-thresholding operation, with a threshold selected such that only the T largest entries in each column of $X+U$ are kept.

III. Experimental Results

To show the advantage of the proposed algorithm, we have performed simulations on a synthetic example. Our ADMM pursuit is compared with the graph regularized sparse coding method by Zheng et al. [7] in representing noisy signals over a known dictionary.

The results presented in Figure 1 clearly demonstrate that the ADMM approach, while being simple and efficient, is advantageous in the achieved representation errors for all the evaluated sparsity levels. In terms of runtime, the two methods are comparable when a very small number of atoms is used, and the ADMM approach is otherwise faster and displays more stable runtimes. This stems from the fact that in our approach the entire matrix X is obtained simultaneously, as opposed to the coordinate descent approach that requires more iterations to converge as the cardinality increases.

Integrating the proposed ADMM pursuit in a dictionary learning framework and evaluating it on real applications, in unsupervised as well as supervised settings, further demonstrates the efficiency of this method.

Algorithm 1 ADMM Pursuit for Manifold Regularized Sparse Coding

Initialize:

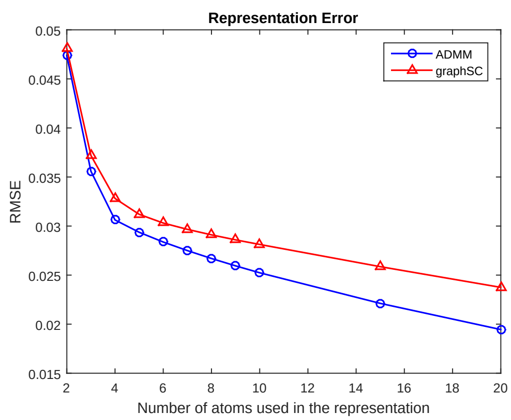
$$X^{(0)} = \arg\min_{X} \|Y - DX\|_F^2 + \beta \sum_i \|x_i\|_1$$

$$Z^{(0)} = X^{(0)} \ , \ U^{(0)} = 0.$$

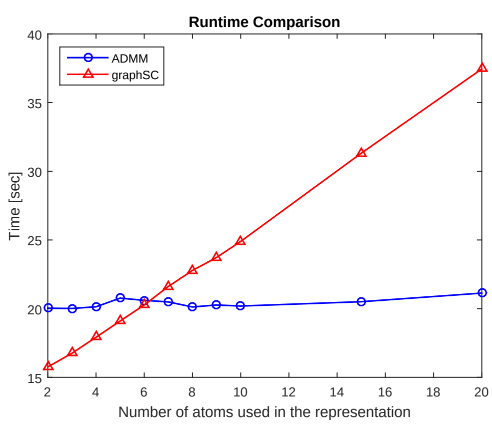
Iterate: for $k = 1, 2, \ldots$

- Update $X^{(k)}$ as the solution of
$$(D^T D + \rho I)X + \gamma X L = D^T Y + \rho \left(Z^{(k-1)} - U^{(k-1)} \right)$$
- Update $Z^{(k)} = \mathcal{P}_{\frac{\beta}{2\rho}} \left(X^{(k)} + U^{(k-1)} \right)$
- Update $U^{(k)} = U^{(k-1)} + X^{(k)} - Z^{(k)}$

Output: The desired result is $Z^{(k)}$.



(a)



(b)

Fig. 1: Evaluation results for two graph regularized pursuit methods: the proposed ADMM solution and the graph regularized sparse coding (graphSC) of [7], in terms of (a) representation error, (b) runtime. For clarity, the evaluated values of β were translated to direct cardinality levels of the representations.

Acknowledgment

The research leading to these results has received funding from the European Research Council under European Union's Seventh Framework Program, ERC Grant agreement no. 320649.

References

[1] M. Elad and M. Aharon, "Image Denoising via Sparse and Redundant Representations Over Learned Dictionaries," *IEEE Trans. Image Proc.*, vol. 15, no. 12, pp. 3736–3745, Dec 2006.

[2] J. Mairal, M. Elad, and G. Sapiro, "Sparse Representation for Color Image Restoration," *IEEE Trans. Img. Proc.*, vol. 17, no. 1, pp. 53–69, 2008. [Online]. Available: http://dx.doi.org/10.1109/TIP.2007.911828

[3] K. Huang and S. Aviyente, "Sparse Representation for Signal Classification," in *Adv. Neural Inf. Process. Syst.*, vol. 19, 2007, pp. 609–616.

[4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, 2009.

[5] K. Kavukcuoglu, M. Ranzato, R. Fergus, and Y. Le-Cun, "Learning Invariant Features Through Topographic Filter Maps," in *CVPR*. IEEE, 2009, pp. 1605–1612.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-Local Sparse Models for Image Restoration," in *ICCV*. IEEE, 2009, pp. 2272–2279.

[7] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph Regularized Sparse Coding for Image Representation," *IEEE Trans. Img. Proc.*, vol. 20, no. 5, pp. 1327–1336, May 2011.

[8] S. Gao, I.-H. Tsang, L.-T. Chia, and P. Zhao, "Local Features Are Not Lonely–Laplacian Sparse Coding for Image Classification," in *CVPR*. IEEE, 2010, pp. 3555–3561.

[9] K. Ramamurthy, J. Thiagarajan, P. Sattigeri, and A. Spanias, "Learning Dictionaries with Graph Embedding Constraints," in *Signals, Systems and Computers (ASILOMAR)*, Nov. 2012, pp. 1974–1978.

[10] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[11] J. Sylvester, "Sur l'equations en matrices $PX = XQ$," *Comptes Rendus Acad. Sci. Paris*, vol. 99, no. 2, pp. 67–71,115–116, 1884.

[12] R. Bhatia, *Matrix Analysis*. Springer-Verlag, New York, 1997.

[13] R. H. Bartels and G. W. Stewart, "Solution of the Matrix Equation $AX + XB = C$," *Comm. ACM*, vol. 15, no. 9, pp. 820–826, Sep. 1972.

[14] G. Golub, S. Nash, and C. Van Loan, "A Hessenberg-Schur Method for the Problem $AX + XB = C$," *IEEE Trans. Auto. Control*, vol. 24, no. 6, pp. 909–913, Dec 1979.

[15] Y. Yankelevsky and M. Elad, "Dual Graph Regularized Dictionary Learning," *IEEE Trans. Sig. Inf. Proc. over Net.*, vol. 2, no. 4, pp. 611–624, Dec. 2016.