

Application of Linearized Alternating Direction Multiplier Method in Dictionary Learning

Xiaoli Yu

School of Science, School of Science and Technology, Shanghai University of Technology, Shanghai, China

Email: 13075511183@163.com

How to cite this paper: Yu, X.L. (2019) Application of Linearized Alternating Direction Multiplier Method in Dictionary Learning. *Journal of Applied Mathematics and Physics*, 7, 138-147.
<https://doi.org/10.4236/jamp.2019.71012>

Received: November 16, 2018

Accepted: January 15, 2019

Published: January 18, 2019

Copyright © 2019 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The Alternating Direction Multiplier Method (ADMM) is widely used in various fields, and different variables are customized in the literature for different application scenarios [1] [2] [3] [4]. Among them, the linearized alternating direction multiplier method (LADMM) has received extensive attention because of its effectiveness and ease of implementation. This paper mainly discusses the application of ADMM in dictionary learning (non-convex problem). Many numerical experiments show that to achieve higher convergence accuracy, the convergence speed of ADMM is slower, especially near the optimal solution. Therefore, we introduce the linearized alternating direction multiplier method (LADMM) to accelerate the convergence speed of ADMM. Specifically, the problem is solved by linearizing the quadratic term of the subproblem, and the convergence of the algorithm is proved. Finally, there is a brief summary of the full text.

Keywords

Alternating Direction Multiplier Method, Dictionary Learning, Linearized Alternating Direction Multiplier, Non-Convex Optimization, Convergence

1. Introduction

With the development of technology, data collection and processing have become easier, and many areas involve high-dimensional data issues, such as information technology, economic finance, and data modeling. Faced with such huge data, many researchers have proposed different solutions, and compressed sensing and sparsity has become an effective algorithm, because sparsity reduces the dimensionality of data in a certain sense and alternates direction multiplier method (ADMM) [5]. It is a typical idea of using divide-and-conquer, which is to transform the original high-dimensional problem into two or more low-dimensional problem-solving algorithms, which is in line with the processing requirements of

big data. However, the traditional ADMM is prone to the local best of the problem. The linear model has a simple structure, which is relatively basic, easy to handle, and widely used. There are many phenomena in real life that can be approximated by a linear model, for example, the relationship between per capita disposable income and consumer spending. Generally speaking, the higher the per capita disposable income X is, the higher the corresponding consumption expenditure Y is. The main advantage of the new algorithm is that it can make the sub-problem have a display solution and be easier to solve, which is of great significance in the application of many problems.

2. Introduction to the Method

The ADMM algorithm was first proposed by Gabay, Meicher and Glowinski in the mid-1970 [6] [7] [8]. A similar idea originated in the mid-1950s. A large number of articles analyzed the nature of this method, but ADMM was used to solve the problem of partial differential equations. Now ADMM is mainly used to solve the optimization problem with separable variables, which solves the problem that the augmented Lagrangian algorithm with good properties can't solve. It can be parallelized, which speeds up the solution. The convergence and convergence rate of ADMM for convex optimization problems with two separable variables. Although there is a mature theoretical analysis, the convergence problem of convex optimization problems extended to more than three separable variables has not been improved in a good solution. Then ADMM is also a public problem for non-convex optimization problems. There have been many applications showing the effectiveness of ADMM for non-convex problems. Can ADMM be applied to more optimization problems and more non-convex optimization problems? What is the effect? This article will introduce the application of ADMM in non-convex optimization problems.

First consider the convex optimization problem with equality constraints

$$\min f(x) \quad \text{s.t. } Ax = b \quad (1)$$

where $x \in R^n$, $A \in R^{m \times n}$, $f: R^n \rightarrow R$ is a convex function.

Firstly, an optimization algorithm with good properties is introduced, which augments the Lagrangian multiplier method. The augmented Lagrangian function is defined as:

$$L_\rho(x, \lambda) = f(x) - \lambda^T (Ax - b) + (\rho/2) \|Ax - b\|_2^2, \quad (2)$$

where $\rho > 0$ is called the penalty parameter. when $\rho = 0$, L_0 is the Lagrangian function. The iterative steps of the augmented Lagrangian multiplier method are:

$$\begin{aligned} x^{(k+1)} &:= \arg \min L_\rho(x, \lambda^k) \\ \lambda^{k+1} &:= \lambda^k - \rho(Ax^{k+1} - b) \end{aligned} \quad (3)$$

where λ is the Lagrangian multiplier, *i.e.* the dual variable.

The advantage of this algorithm is that the convergence of the iterative se-

quence can be guaranteed without too strong conditions. For example, for the penalty parameter, it is not required to increase to infinity in the iterative process, and a fixed value can be taken. But the disadvantage of this algorithm is that when the objective function is separable, the model becomes:

$$\begin{aligned} \min & f(x) + g(y) \\ \text{s.t.} & Ax + By = b \end{aligned} \quad (4)$$

where g is also a convex function. In the x iteration step, the augmented Lagrangian function L_ρ is inseparable, and the discrete variables cannot be solved in parallel for the x subproblem. This leads to the ADMM algorithm we will discuss in the next section. The Alternating Direction Method (ADMM) is mainly used to solve the optimization problem with separable variables like (4), where $x \in R^n, y \in R^m, A \in R^{p \times n}, B \in R^{p \times m}, b \in R^p$. Let's assume that both f and g are convex functions, and then make other assumptions. Similar to the definition in the previous section, the augmented Lagrangian penalty function of (4) is:

$$L_\rho(x, y, \lambda) = f(x) + g(y) - \lambda^T (Ax + By - b) + (\rho/2) \|Ax + By - b\|_2^2 \quad (5)$$

The steps of the ADMM algorithm iteration are as follows:

$$\begin{aligned} x^{k+1} &:= \arg \min L_\rho(x, y^k, \lambda^k) \\ y^{k+1} &:= \arg \min L_\rho(x^{k+1}, y, \lambda^k) \\ \lambda^{k+1} &:= \lambda^k - \rho(Ax^{(k+1)} + By^{(k+1)} - b) \end{aligned} \quad (6)$$

where $\rho > 0$. The similarity between the algorithm and the augmented Lagrangian multiplier method is to iteratively solve the variables x and y and then iteratively solve the dual variables.

If the augmented Lagrangian multiplier method is used for iterative solution:

$$\begin{aligned} (x^{(k+1)}, y^{(k+1)}) &:= \arg \min L_\rho(x, y, \lambda^{(k)}) \\ \lambda^{k+1} &:= \lambda^k - \rho(Ax^{(k+1)} + By^{(k+1)} - b) \end{aligned} \quad (7)$$

As mentioned in the previous section, you can see that the augmented Lagrangian multiplier method deals with two separate variables at the same time, and ADMM alternates the variables, which is the origin of the algorithm name. It can be considered that this is the use of Gauss-Seidel iterations on two variables. For details, please refer to. It is obvious from the algorithm framework that the ADMM algorithm is more suitable for solving the problem of having separate variables because the objective functions f and g are also separated.

To get a simpler form of ADMM, normalize the dual variable so that $\mu = (1/\rho)\lambda$. Then the ADMM iteration becomes:

$$\begin{aligned} x^{(k+1)} &:= \arg \min \left(f(x) + (\rho/2) \|Ax + By^{(k)} - b + \mu^{(k)}\|_2^2 \right) \\ y^{(k+1)} &:= \arg \min \left(g(y) + (\rho/2) \|Ax^{k+1} + By - b + \mu^{(k)}\|_2^2 \right) \\ \mu^{(k+1)} &:= \mu^{(k)} - (Ax^{(k+1)} + By^{(k+1)} - b) \end{aligned} \quad (8)$$

ADMM convergence: Regarding the convergence of ADMM, please refer to

the literature.

3. Application of ADMM in Dictionary Learning

As we all know, the alternating direction method (ADMM) is one of the effective algorithms for solving large-scale sparse optimization problems. It is solved by splitting the problem into a number of low-dimensional sub-problems by augmented Lagrangian function construction. In recent years, a large number of working signals have pointed to the sparse expression of signals. Sparse expression refers to the use of a $D \in R^{m \times n}$ ($m \ll n$) dictionary, the dictionary contains n Signal atoms $\{d_j\}_{j=1}^n$. A signal $y \in R^m$ can be expressed as a sparse linear representation of these signal atoms. In fact, the so-called sparse means that the number of non-zero coefficients is much smaller than that of n . Such a sparse representation may be a determined $y = Dx$ or an approximate representation with an error term $\|y - Dx\|_p \leq \varepsilon$. The vector $x \in R^n$ is the signal y sparse expression coefficient. In practice, p often takes a value of 1, 2, or ∞ .

If $m < n$ and the dictionary D is full rank, then the underdetermined system of the problem has an infinite number of solutions, and the solution using the least non-zero coefficient is one of them, and is the solution we hope to find. Sparse expression is expressed as a mathematical expression

$$(P_0) \min \|x\|_0 \quad \text{subject to } y = Dx \quad (9)$$

Or

$$(P_0^\epsilon) \min \|x\|_0 \quad \text{subject to } \|y - Dx\|_2 \leq \epsilon \quad (10)$$

where $\|\cdot\|_0$ is l_0 -module, which means that the corresponding vector takes a non-zero quantity.

Dictionary Design

Learn the dictionary based on the signal set. First given a data set $Y = \{y_i\}_{i=1}^L$, assuming that there is a dictionary D so that for a given signal can be represented as a sparse representation of the dictionary, *i.e.* for a given signal y_i , the model (P_0) Or (P_1) can find the sparse coefficient x_i . The question then is how to find such a dictionary D . Detailed reference can be found in the literature [9].

The model of the problem can be written as:

$$\min_{D,X} \left\{ \|Y - Dx\|_F^2 \right\} \quad \text{s.t. } \|x_i\|_0 \leq \tau_0, i = 1, \dots, L \quad (11)$$

where τ_0 is the upper bound of the coefficient sparsity, x_i is the i th column of the coefficient matrix X , and $\|\cdot\|_F^2$ is the Frobenius norm of the matrix, *i.e.* the sum of the squares of the elements of the matrix.

Another model for dictionary learning is corresponding to the above model.

$$\min_{D,X} \sum_{i=1}^L \|x_i\|_0 \quad \text{s.t. } \left\{ \|Y - Dx\|_F^2 \right\} \leq \epsilon \quad (12)$$

ϵ is a fixed error value.

Before applying ADMM, first make some transformations to the model, let

$Z = DX$, then the model becomes:

$$\min_{D, X, Z} \left\{ \|Y - DX\|_F^2 \right\} \quad \text{s.t. } Z = DX, \|x_i\|_0 \leq \tau_0, i = 1, \dots, L \quad (13)$$

Then the augmented Lagrangian function of the problem is:

$$L_\beta(D, X, Z, \Lambda) := \|Y - DX\|_F^2 + \sum_{i=1}^L \langle \Lambda_i, (Z - DX)_i \rangle + \frac{\beta}{2} \|Z - DX\|_F^2 \quad (14)$$

where Λ is the Lagrange multiplier matrix and Λ_i is the i th column of Λ .

Using the ADMM algorithm in the above model, there is a X subproblem

$$\min_X \sum_{i=1}^L \langle \Lambda_i, (Z - DX)_i \rangle + \frac{\beta}{2} \|Z - DX\|_F^2 \quad \text{s.t. } \|x_i\|_0 \leq \tau_0, i = 1, \dots, L \quad (15)$$

Equivalent to

$$\min_X \frac{\beta}{2} \|Z + \Lambda/\beta - DX\|_F^2 \quad \text{s.t. } \|x_i\|_0 \leq \tau_0, i = 1, \dots, L \quad (16)$$

The Z subproblem is

$$\min_Z \|Y - Z\|_F^2 + \sum_{i=1}^L \langle \Lambda_i, (Z - DX)_i \rangle + \frac{\beta}{2} \|Z - DX\|_F^2 \quad (17)$$

This sub-question has a solution.

$$Z = (\beta DX + 2Y - \Lambda)/(2 + \beta) \quad (18)$$

D sub-problem is

$$\min_D \frac{\beta}{2} \|Z + \Lambda/\beta - DX\|_F^2 \quad (19)$$

Λ updated to

$$\Lambda^{(k+1)} = \Lambda^{(k)} + \gamma\beta(Z - DX) \quad (20)$$

But ADL (ADMM for Dictionary Learning) is prone to the local best of the problem. Using linearization techniques, we extended LADMM to solve the problem of ADL local straits and proved the convergence of the algorithm. Numerical experiments are used to illustrate the effectiveness of the proposed algorithm.

4. Application of Linearized ADMM in Dictionary Learning

In order to apply ADMM, we can rewrite (11) into the following form

$$\min_{D, X} \sum_{i=1}^L \|x_i\|_0 \quad \text{s.t. } Y = DX \quad (21)$$

Then the augmented Lagrangian function of the model (20) is

$$L_\mu(X, Y, \lambda) = \sum_{i=1}^L \|x_i\|_0 - \lambda^T(Y - DX) + \frac{\mu}{2} \|Y - DX\|_2^2 \quad (22)$$

The iterative method of ADMM is:

$$\begin{cases} X^{k+1} = \arg \min L_\mu(X, Y^k, \lambda^k) \\ Y^{k+1} = \arg \min L_\mu(X^{k+1}, Y, \lambda^k) \\ \lambda^{k+1} = \lambda^k - \mu(Y^{k+1} - DX^{k+1}) \end{cases} \quad (23)$$

Now, we solve the subproblem in (22). First we solve the X -sub problem.

$$\begin{aligned}
X^{k+1} &= \arg \min \left\{ \sum_{i=1}^L \|x_i\|_0 - (\lambda^k)^T (Y^k - DX) + \frac{\mu}{2} \|Y^k - DX\|_2^2 \right\} \\
&= \arg \min \left\{ \sum_{i=1}^L \|x_i\|_0 - (\lambda^k)^T (Y^k - DX) + \frac{\mu}{2} \|Y^k - DX\|_2^2 \right\} \\
&= \sum_{i=1}^L \|x_i\|_0 + \frac{\mu}{2} \|Y^k - DX - \lambda^k / \mu\|_2^2
\end{aligned} \tag{24}$$

Because of the non-identity of matrix D , this subproblem does not show a solution. Inspired by [9], we linearize this quadratic term $\frac{1}{2} \|Y^k - DX - \lambda^k / \mu\|_2^2$ as

$$D^T (DX^k - Y^k - \lambda^k / \mu)^T (X - X^k) + \frac{\rho}{2} \|X - X^k\|_2^2 \tag{25}$$

The parameter $\rho > 0$ controls the degree of approximation of X and X^k , then we solve the following problem and use the solution of this problem to approximate the solution of the subproblem generated by ADMM.

$$\begin{aligned}
X^k &= \arg \min \left\{ \sum_{i=1}^L \|x_i\|_0 + \frac{\mu\rho}{2} \|X - X^k\|_2^2 \right. \\
&\quad \left. + \mu \left[D^T (DX^k - Y^k - \lambda^k / \mu)^T (X - X^k) \right] \right\} \\
&= \arg \min \left\{ \sum_{i=1}^L \|x_i\|_0 + \frac{\mu\rho}{2} \|X - X^k + D^T (DX^k - Y^k + \lambda^k / \mu) / \rho\|_2^2 \right\}
\end{aligned} \tag{26}$$

For the above problem, it is known from [9] medium (11)

$$\begin{aligned}
(x_i)^{k+1} &= \arg \min \left\{ \sum_{i=1}^L \|x_i\|_0 + \frac{\mu\rho}{2} \left\| x_i - \left(X^k - \frac{D^T (DX^k - Y^k - \lambda^k / \mu)}{\rho} \right) \right\|_2^2 \right\} \\
&= \text{shrink} \left(\left(X^k - D^T (DX^k - Y^k - \lambda^k / \mu) / \rho \right)_i, \frac{\lambda}{\mu\rho} \right)
\end{aligned} \tag{27}$$

Furthermore, for the Y subproblem, the Equation (10) in [9] shows that the display solution is

$$Y^{k+1} = \text{shrink}_{1,2} \left(DX^{k+1} - \frac{\lambda^k}{\mu}, \frac{1}{\mu} \right) \tag{28}$$

As can be seen from the above discussion, the LADMM iterative algorithm can be described by the following table.

Input: X, Y and tol. Choose $\lambda_i > 0, \mu > 0$ and $\rho > \rho(X^T X + \lambda_i^2 I_p)$, where $\rho(\cdot)$ denotes the spectral radius, select $(X^0, Y^0, \lambda^0) = (0, 0, 0)$ for $k = 1, 2, \dots, N$

‘do

 Compute X^{k+1} by (27).

 Compute Y^{k+1} by (28).

 end

 Update $\lambda^{k+1} = \lambda^k - \mu(Y^{k+1} - DX^{k+1})$;

end

Output: (X^N, Y^N, λ^N) as an approximate solution of (11).

Convergence Proof

In this section, we will demonstrate that the LADMM algorithm is convergent. I_p is the unit matrix of $p \times p$. Let $S = R^p \times R^{n+p} \times R^{n+p}$ and $\omega = (\beta^T, \gamma^T, \alpha^T)^T$, $f(\beta) \in \partial(\|\beta\|_{2,1})$, $g(\gamma) \in \partial(\|\gamma\|_1)$. Solving the model is equivalent to finding $(\beta^*, \gamma^*, \alpha^*)$ that satisfies the KKT condition of the model, *i.e.*

$$\begin{cases} \lambda_2 f(\beta^*) - \tilde{X}^T \alpha^* = 0 \\ g(\gamma^*) + \alpha^* = 0 \\ \tilde{X} \beta^* - \tilde{y} - \gamma^* = 0 \end{cases} \quad (29)$$

Let us remember that the set of elements that satisfy the above formula in S is S^* . The KKT condition of the above formula can be written as the form of variational inequality (VI) as:

$$(\omega - \omega^*)^T F(\omega^*) \geq 0, \quad \forall \omega \in S, \quad (30)$$

where

$$F(\omega) = \begin{cases} \lambda_2 f(\beta) - \tilde{X}^T \alpha \\ g(\gamma) + \alpha \\ \tilde{X} \beta - \tilde{y} - \gamma \end{cases} \quad (31)$$

In order to prove these conclusions, as well as the proof of convergence of LADMM, need to introduce some lemma. For details, please refer to the literature [9].

5. Numerical Experiments

In this chapter, we will discuss the application of the algorithm in image deblurring to prove the effectiveness of the algorithm. All experiments were carried out on a four-core notebook computer with Intel(R) Core(TM) i5-7200UCPU @ 2.50 GHz and 4 GB memory. Procedures for this experiment, pictures are referenced [10].

Figure 1 shows the application of ADMM algorithm in image deblurring, which has motion blur and Gaussian blur respectively. (“motion”, 35, 50), (“gaussian”, 20, 20).

The noise levels are $\delta = 0.256, 0.511$ respectively. For comparison, we also include the results of FTVd v4.1 in [11], which is the most advanced image deblurring algorithm. It can be seen from the pictures that our proposed algorithm and FTVd algorithm have the same quality as PSNR (**Figure 2**), and our algorithm does not need regularization operator.

$$\text{PSNR} = 10 \cdot \log_{10} \frac{255^2}{\text{MSE}} [\text{dB}]$$

MSE represents the average square error of each pixel.

6. Conclusions

In this paper, we propose a linearized alternating direction multiplier method



Figure 1. ADMM algorithm for deblurring.

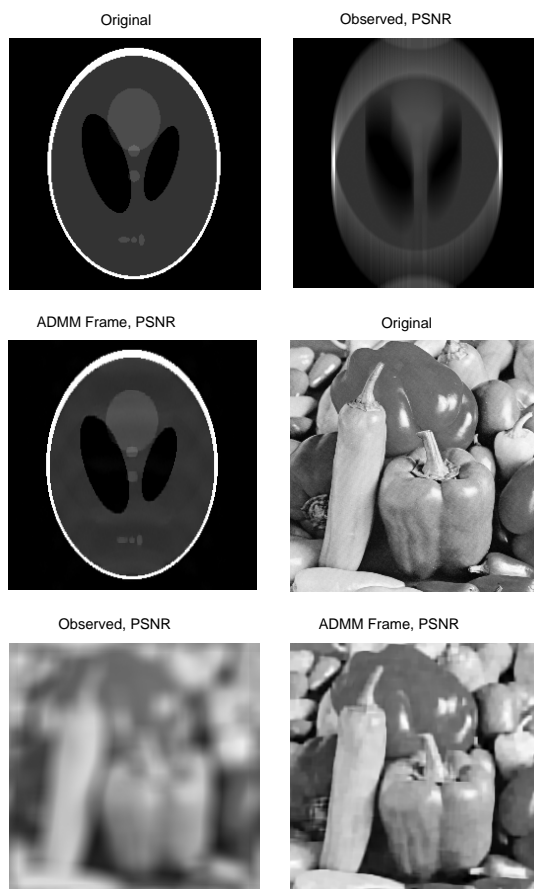


Figure 2. LADMM algorithm for deblurring.

(LADMM) to solve the problem of rapid convergence of the dictionary model that is easy to fall into the local optimal solution. This model combines the ad-

vantages of linearization features to make sub-problems easier to solve. When the algorithm is applied to the problem of image reconstruction, we use the quadratic term of the linearized sub-problem to make the sub-problem easier to solve.

Moreover, in the case of similar PSNR, the algorithm does not need regularization operator.

In addition, we analyze the convergence of the LADMM algorithm. Our next step is to generalize the model to more non-convex problems (non-negative matrix factorization problems) and use practical problems.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Boyd, S., Parikh, N., Chu, E., Peleato, B. and Eckstein, J. (2010) Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. *Foundations and Trends® in Machine Learning*, **3**, 1-122. <https://doi.org/10.1561/22000000016>
- [2] Nadai E. (2011) Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer, Berlin.
- [3] Wu, Z.M. (2016) Linearized Alternating Direction Method for Solving Separable Convex Optimization Problems. <https://wenku.baidu.com/view/7aff15227ed5360cba1aa8114431b90d6c8589e3.html>
- [4] Fu, X.L., He, B.H., Wang, X.F., *et al.* (2014) Block-Wise Alternating Direction Method of Multipliers with Gaussian Back Substitution for Multiple-Block Convex Programming. 1-37. http://xueshu.baidu.com/usercenter/paper/show?paperid=4cb7be88f0f61a585690896f1da8bf31&site=xueshu_se
- [5] Wang, H.F. and Kong, L.C. (2017) The Linearized Alternating Direction Method of Multipliers for Sparse Group LAD Model. Beijing Jiaotong University, Beijing.
- [6] Eckstein, J. (2016) Approximate ADMM Algorithms Derived from Lagrangian Splitting. *Computational Optimization and Applications*, **68**, 363-405. <https://doi.org/10.1007/s10589-017-9911-z>
- [7] Haubruge, S., Nguyen, V.H. and Strodiot, J.J. (1998) Convergence Analysis and Applications of the Glowinski—Le Tallec Splitting Method for Finding a Zero of the Sum of Two Maximal Monotone Operators. *Journal of Optimization Theory and Applications*, **97**, 645-673.
- [8] Glowinski, R. and Marroco, A. (1975) Sur l'approximation, par éléments finis d'ordre un, et la résolution, par Pénalisation-Dualité d'une classe de problèmes de Dirichlet non linéaires. *Journal of Equine Veterinary Science*, **2**, 41-76.
- [9] Wang, H.F. (2017) Linearized Multiplier Alternating Direction Method for Solving Least Degree Model of Sparse Group. <https://wenku.baidu.com/view/f25e8ff1d5d8d15abe23482fb4daa58da0111c9c.html>
- [10] Jiao, Y.L., Jin, Q.N., Lu, X.L., *et al.* (2016) Alternating Direction Method of Multipliers for Linear Inverse Problems. *SIAM Journal on Numerical Analysis*, **54**, 2114-2137. <https://doi.org/10.1137/15M1029308>

- [11] Wang, Y.L., Yang, J.F., Yin, W.T., *et al.* (2016) A New Alternating Minimization Algorithm for Total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, **1**, 248–272.