Data Reduction Algorithm Using Nonnegative Matrix Factorization with Nonlinear Constraints

PasukatSembiring

1Department of Mathematics, University of Sumatera Utara, Medan, Indonesia
Email: pasukat_sembiring366@gmail.com

Abstract—Processing of data with very large dimensions has been a hot topic in recent decades. Various techniques have been proposed in order to execute the desired information or structure. Non- Negative Matrix Factorization (NMF) based on non-negatives data has become one of the popular methods for shrinking dimensions. The main strength of this method is non-negative object, the object model by a combination of some basic non-negative parts, so as to provide a physical interpretation of the object construction. The NMF is a dimension reduction method that has been used widely for numerous applications including computer vision, text mining, pattern recognitions, and bioinformatics. Mathematical formulation for NMF did not appear as a convex optimization problem and various types of algorithms have been proposed to solve the problem. The Framework of Alternative Nonnegative Least Square (ANLS) are the coordinates of the block formulation approaches that have been proven reliable theoretically and empirically efficient. This paper proposes a new algorithm to solve NMF problem based on the framework of ANLS. This algorithm inherits the convergence property of the ANLS framework to nonlinear constraints NMF formulations.

Keyword: Dimensional Reduction, Nonnegative Matrix Factorization, Sparse Matrix, Nonlinear Constraints

1. Introduction
In every second in the modern era, a lot of data sets are being generated. The number of people online write their blogs, their homepage design and disseminate their experiences through digital support, videos, images and others. Imagine also the data formed from living organisms and gene research data obtained from space or from our earth. Data of transaction’s through e-banking and others. Be useful if the data has been processed. Associated with the boom in the amount of data so rapidly, there are several approaches for data processing, by using classical methods application, designing more powerful computing structures such as: distributed computing, multi-core processors, super computers and others. However, the amount of data growth and complexities tendency exceeds the increase computational capabilities [1]. One of very popular approach called model of reduction that attempts to reduce the complexity while maintaining the assumption of issues (preliminary data). The use of an appropriate model can save the time. In addition, different data types require different models to capture the data comprehension. Of course, a model considered appropriate so that the model can survive better
met. An ideal model may not exist. For example, using sub-dominant space with the Singular Value Decomposition (SVD) has been proposed as the best model for reducing the complexity of data and complicated system [10]. This model offers the smallest error with the same complexity shrinkage, compared with other models. Partial-based analysis also referred to as additive model in the absence of the operating margin in the model. This is related to the formation of an object:

\[
\text{Object}_i = \text{Part}_1(b_{i1}) \text{ with } \text{Part}_2(b_{i2}) \ldots ,
\]

Where, \( b_{ij} = \) present if part \( i \) is present in object \( j \) or, \( b_{ij} = \) absent if part \( i \) is absent from object \( j \). Their operating margin, which \( b_{ij} < 0 \) for some \( ij \) result in each section can be interpreted as positive or negative quantity. If each part included in a vector space, this is true because orientation does not change the room establishment. But for other data types, such as: concentration, relative temperature, opportunities, and other sound spectrum - more negative quantity does not appear. Decomposition of non-negative objects with common methods such as SVD significantly alters the physical interpretation of the data. Other analyzes such as Principle Components Analysis (PCA) requires some features where the object is non-negative because nature is not owned, such sum to zero, orthogonality and others. So much so that became the main reference of this dissertation research. To focus on the data nonnegative through the study of non-negative matrix factorization (FMN). This technique explores for applying the non-negative object \( A \) by two other non-negative matrix \( W \) and \( H \).

The rest of the paper is organized as follows. The concepts of Nonnegative Matrix Factorization (NMF) and related work are explained in section II. Proposed embedding and extraction algorithms are explained in section II. Experimental results are presented in section III. Concluding remarks are given in section IV.

2. Nonnegative Matrix Factorization

Nonnegative Matrix Factorization (NMF), was originally known as the factorization rank non-negative or matrix factorizations positive [11] have attracted much attention as a method of dimension reduction in machine learning and data mining [10,11]. NMF is used for high dimensional data in which each nonnegative valuable element. Due to the nature of this non-negative factor of the approach is much smaller give an interpretation. Each item of the data can be explained by a combination of linear additives from line items that are physically meaningful [10]. Various application of NMF has been reported in the field including text mining [7], bioinformatics [16].

Mathematically NMF can be formulated in the following way. Given an input matrix \( A \in \mathbb{R}^{m \times n} \), in which each element nonnegative, an integer \( k < \min \{m, n\} \) NMF aims to acquire two vectors \( W \in \mathbb{R}^{m \times k} \) and \( H \in \mathbb{R}^{k \times n} \) with non-negative elements, so

\[
A \approx WH
\]  

(2.1)

Measurement approach in Eq. (2.1) can be done in several ways, among other things, non Frobenius, divergence Kullbach - Leibber [10] and the divergence Bregaran [19]. The research was wearing a size selection of the most secure, ie non Frobenius. So, the factor \( W \) and \( H \) are obtained by solving the following optimization.

\[
\min_W \sum_{i=1}^{m} \sum_{j=1}^{n} f(W,H) = \frac{1}{2} ||A - WH||^2_F
\]  

(2.2)

Where inequality \( W \geq 0 \) and \( H \geq 0 \) gives the sense that all of the elements \( W \) and \( H \) nonnegative. The issue in Eq. (2) is difficult. Various algorithms have been developed to solve Equation (2). At the first Paatero and Tappen [11] proposed NMF (or, more precisely positive matrix factorization) but their algorithm is not exactly related to the non-negative constraint. The most popular approach to solve Eq. (2) is the multiplicative update algorithm of Lee and Seung [10]. At each of iteration of this method, the elements \( W \) and \( H \) multiplied by a certain factor. Because the zero element is not in
update, all components of W and H positive for all iterations. However, some issues about the performance [20], as well as the issue of convergence [21] reported. In [18], a simple algorithm, that resolve issues without the constraints of least squares at each iteration, has been used but the algorithm also has difficulty with convergence. Furthermore, some algorithms that are based on the framework of alternative non-negative least squares (ANLS) filed with the well's Performance [18].

Non-negative matrix factorization first introduced by Paatero and Tapper in [11]. But this term gained popularity thanks to Lee and Seung writing [10]. They propose that non-negativity important in human perception and also provide two simple algorithms for determining the representation of non-negative for non-negative data. Known $m \times n$ non-negative matrix A (ie. $A_{ij} \geq 0$) and reduced rank $r \leq \min(m,n)$, non-negative matrix factorization problem consists of the determination of the two non-negative matrix $U \in \mathbb{R}^{m \times r}$ and $V \in \mathbb{R}^{n \times r}$ that approximates the matrix A, that is:

$$A \approx UV^T.$$  

(2.3)

By paying attention to the columns of $A$, seen that each of these columns is approximated by a combination of non-negative $r$ basis vectors which are the columns of $U$. Could be considered, the columns of $U$ as a base of $U$, which is entirely in the non-negative. Each column $A$ is approximated by an element of $U$, usually the element closest to the column. Also can exchange the role of $U$ and $V$

to affirm that each row of $A$ is approximated by an element of $V$, usually the element closest to the line, where $V$ is generated by the columns of $V$. Where,

$$A_i \approx \sum_{j=1}^{r} V_{i,j} U_{j, \cdot}.$$  

(2.4)

There are several ways to determine the difference between the data matrix $A$ and matrix models $UV^T$. But the most widely used measure is the Frobenius norm:

$$F(A,UV^T) = \frac{1}{2} \|A - UV^T\|_F^2 = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} (A_{ij} - [UV^T]_{ij})^2.$$  

(2.5)

This is also referred to as distance Euclid (Euclidean Distance).

Suppose that $U$ remains, the function $F(U, V) = \frac{1}{2} \|A - UV^T\|_F^2$ can be viewed as the composition of the Frobenius norm and linear transformation of $V$. Therefore, $F$ is convex in $V$. Similarly, if $V$ is fixed, $F$ is convex on $U$.

In this study, non-negative matrix factorization problem will be studied with Frobenius norm. The main issues are as follows:

$$\min_{U \in \mathbb{R}^{m \times r}, V \in \mathbb{R}^{n \times r}} \frac{1}{2} \|A - UV^T\|_F^2.$$  

(2.6)

The problem is Nonnegative Matrix Factorization. Finish it, when $m \times n$ non-negative matrix $A$ and integer $r \leq \min (m, n)$. Where $r$ is called the rank reduced. Henceforth, $m$ and $n$ will be used to denote the size of the matrix $A$ and $r$ is the rank reduced from a factorization. Nonnegative matrix factorization rewritten as a nonlinear optimization problem:

$$\min_{U \geq 0, V \geq 0} \frac{1}{2} \|A - UV^T\|_F^2.$$  

(2.7)

Lagrange function associated is:

$$L(U,V,\mu,\nu) = \frac{1}{2} \|A - UV^T\|_F^2 - \mu \circ U - \nu \circ V,$$  

(2.8)
Where $\mu$ and $\nu$ are two equal-sized matrices, $U$ and $V$ containing multipliers Lagrange associated with restrictions nonnegativity $U_{ij} \geq 0$ dan $V_{ij} \geq 0$. Then the terms of Karush-Kuhn-Tucker, for nonnegative matrix factorization problem is, If $(U, V)$ is a local minimum, then there $\mu_{ij} \geq 0$ dan $\nu_{ij} \geq 0$ such that:

$$U \geq 0, \quad V \geq 0,$$

(2.9)

$$\nabla L_U = 0, \quad \nabla L_V = 0,$$

(2.10)

$$\mu \circ U = 0, \quad \nu \circ V = 0.$$  

(2.11)

By developing (2.10) were obtained:

$$AV - UV^T V - \mu = 0, \quad A^T U - VU^T U - \nu = 0 \quad \mbox{or} \quad \mu = -(UV^T V - AV), \quad \nu = -(VU^T U - A^T U).$$

By combining this with $\mu_{ij} \geq 0$, $\nu_{ij} \geq 0$ and (2.11) provides the following terms:

$$U \geq 0, \quad V \geq 0,$$

(2.12)

$$VF_U = UV^T V - AV \geq 0, \quad VF_V = VU^T U - A^T U \geq 0,$$

(2.13)

$$U \circ (UV^T V - AV) = 0, \quad V \circ (VU^T U - A^T U) = 0.$$  

(2.14)

Where the Lagrange multiplier is corresponding to the matrix $U$ and $V$ is also the gradient of $F$ to matrix $U$ and $V$. Since the distance Euclidean is not convex in the variables $U$ and $V$, so that the terms of this is only a necessary condition. This is implied because of the existence of a saddle point and maximum

3. Non Linear Programming

In the section, it was found a certain optimization procedure to maximize or minimize a function of $n$ variables. Worse, as noted earlier, this procedure cannot be used to solve problems that are very much exceeds its variable. Here will be presented a few, computationally feasible method to solve certain types of problems, non-linear programming. One problem, in nonlinear programming problem, which is must to maximize or minimize a function $f(\vec{x})$, the constraints of a set of equations or inequality constraints. Where $f(\vec{x})$, or at least one of the functions that appear in this set of constraints or both, is a non-linear function. The Problem of non-linear programming generally can be expressed as follows:

$$\text{Maximize } Z = f(\vec{x})$$  

(3.1)

Constraints $gi(x) \{ s =, \leq \} b(i)$

(3.2)

Weir, one of the three relations $\{ s =, \leq \}$ is designed against each of $m$ constraints, the function $f(\vec{x})$ in equation (3.1) calls the objective function. In connection with the non-negativity constraint ($x_j \geq 0$) on some, or all, of the variables stated, separately or may be deemed to be included in the constraint. There is no known method of determining the global maximum of the problem, non-linear programming in general. Even, if the objective function and constraints in accordance with the properties of certain global maximum sometimes acceptable. For example proves that the global maximum of a function convex against a limited set convex and on, an extreme point of the set convex. Here will be downgraded "Kuhn-Tucker conditions" a set of conditions necessary, to make the maximum of a non-linear problems. These conditions in certain cases also produce global maximum. Kuhn-Tucker conditions are very useful indeed, in generating methods to resolve some types of non-linear problems. It can be seen how the use of Kuhn-Tucker conditions to develop an algorithm using the Simplex method to solve quadratic program. The Problem in the quadratic program is the objective function is quadratic and the constraints are linear

4. Proposed Algorithm
Mathematical formulation of NMF can be expressed in the following way. Known input matrices $A \in \mathbb{R}^{M \times n}$ in which each element and a nonnegative integer $k < \min \{m, n\}$, NMF aims to determine two facts $W \in \mathbb{R}^{m \times k}$ and $H \in \mathbb{R}^{k \times n}$ with nonnegative elements, so:

$$A = WH$$  \hspace{1cm} (4.1)

By using Fortenius norms, facts $W$ and $H$ is obtained by solving the following problems sptiuseri:

$$\min_{N \geq 0, H \geq 0} f(N, H) = \frac{1}{2} \|A - WH\|_F^2$$  \hspace{1cm} (4.2)

Where the inequality $W \geq 0$ and $H \geq 0$ means that all elements of $W$ and $H$ nonnegative.

Before proposing the algorithm to solve Eq.(4.2). This is necessary because the proposed algorithm starts from ANLS. Within the framework of the basic ANLS, the variable directly divided into two groups. Then both of groups updated. This basic framework can be summarized as follows:

1. Initialization $H \in \mathbb{R}^{m \times 1}$ with elements non negative
2. Complete the following problems until termination criteria are met:

$$\min_{W \geq 0} \|WH - A\|_F^2$$  \hspace{1cm} (4.3)

With $H$ fixed, and

$$\min_{H \geq 0} \|WH - A\|_F^2$$  \hspace{1cm} (4.4)

And $W$ fixed

3. Column $W$ is normalized into the unit $L_2$-norm and the corresponding rows of $H$ scaled.

Alternatively, $W$ can be just as initial iteration of data later in the Eq (4.4) and then Eq. (4.3). Note that each sub problems is a question of non-negativity least square (NNLS). Although, the initial problem (4.2) is non-convex, sub problem Eq. (4.3) is convex. NNLS issue in Eq (4.3) and has certain characteristics. NMF is an algorithm to reduce the dimensions. Very large initial dimensions, eg thousands, of small dimensions reduced. Because the matrix $W \in \mathbb{R}^{m \times 1}$ in Eq. (4.4) shape is very long and thin $(m >> k)$ and the coefficient matrix $H^T \in \mathbb{R}^{m \times 1}$ in Eq. (4.3) is also long and thin $(n >> k)$. Because the matrix $W^T$ and $H$ of the variables in Eq. (4.3) and (4.4) each of data and wide. To design the NMF algorithm based on ANLS framework, the need for a method to solve sub problem in Eq. (4.3) and (4.4). The algorithm is a kelaril method to NNLS active set by Lawson and Hanson [9]. The Set method actively seeking optimal set of active and passive by maintaining between two sets of variables. If the variable passive (ie, positive) solution is known, then the issue can be resolved by procedures of NNLS least square has no control over passive variable. The main limitation of this method is that the active set of the set of variables that meet non-negativity maintained precede vector settlement, while ensuring that the objective function decreases in each iteration. Means algorithm is slow when a large number unknown. Pivot major blocks method is proposed to identify the limitations of active set methods.

The first, this paper presents the block principal pivoting method for the NNLS problems with single right-hand side and then proposed the method developed to handle multiple right-hand sides. The NNLS problems with single right-hand side vector formulated as follow:

$$\min_{x \geq 0} \|Cx - b\|_2^2$$  \hspace{1cm} (4.5)

Where, $C \in \mathbb{R}^{p \times q}$, $b \in \mathbb{R}^{q \times 1}$, and $p \geq q$. The sub problems in Eq. (4.3) and (4.4) can be decomposed into several independent instances of Eq.(4.5). Thus, an algorithm for Eq.(4.5) is a basic building block for an algorithm Eq. (4.3) and Eq.(4.4).

The Karush – Kuhn- Tucher optimality condition for Eq.(4.5) is written as follows:
\[ y = C^T C x - C^T b \]  
\[ y \geq 0 \]  
\[ x \geq 0 \]  
\[ x, y = 0, I = 1, \ldots, \infty \]

(4.6) - (4.9)

It is assumed that the matrix \( C \) has full column rank. In this case, the matrix positive \( C^T C \) is positive definite and the problems in Eq. (4.5) is strictly convex. Then, a solution \( x \) that satisfies the conditions in Eq. (4.6) - (4.9) is the optimal solution of Eq. (4.5).

The set of indexes \{ 1, \ldots, \} are divided into two groups \( F \) and \( G \) to \( F \cup G = \{ 1, \ldots, \} \) and \( F \cap C = \emptyset \). Let's, \( X_F, X_G, Y_F \) and \( Y_G \) denote sub matrix \( C \) with corresponding indexes, and suppose \( C_F \) and \( C_G \) stated sub matrix \( C \) with corresponding column index initially: Take \( X_G = 0 \) and \( Y_F = 0 \), all the elements \( X_G \) and \( Y_F \) made zero. Thus, \( x \) and \( y \) have always satisfied the Eq. (4.1) for each price of \( X_F \) and \( Y_G \). Now, calculate \( X_F \) and \( Y_G \) using Eq. (4.6) and then check whether the computed value of \( X_F \) and \( Y_G \) meets the Eq. (4.7) and (4.8). The calculation of the \( X_F \) and \( Y_G \) is done by:

\[ X_F = \arg \min_{X_F} \| C_F X_F - b \|_2^2 \]  
\[ Y_G = C^T_G (C_F X_F - b) \]

(4.10) - (4.11)

Complete the Eq. (4.10) first for \( X_F \) then substitutes the result to the Eq. (4.11). Couples \( (X_F, Y_G) \) are called basic complementary if they are obtained by Eq. (5.10) - (5.11). If the basic complementary settlement \( (X_F, Y_G) \) meets \( X_F \emptyset \) and \( Y_G \emptyset \), then it is called feasible. In this case \( X \) is the optimal solution of the Eq. (4.5) and BerLouthalgorithms. Otherwise, the complementary basic settlement \( (X_F, Y_G) \) is not feasible, and it needs to be updated F and G by exchanging variables which are the Eq. (5.7) and (5.8) do not apply. So, the definition of the following index set:

\[ V = \{ I \in F : x_i < 0 \} \cup \{ I \in G : y_i < 0 \} \]

(4.12)

The variable \( x \) with \( i \in V \) is called an improper variable. Now select an empty subset of \( \tilde{V} \subset V \) then, \( F \) and \( C \) are absolute with the following rules:

\[ F = (F - \tilde{V}) \cup (\tilde{V} \cap C) \]
\[ C = (G - \tilde{V}) \cup (\tilde{V} \cap F) \]

(4.13) - (4.14)

Size \( |\tilde{V}| \) presents how many variables are exchanged per iteration. If \( |\tilde{V}| = 1 \), the algorithm is called the main single pivot algorithm. The algorithm repeats this procedure so that the number of unworthy variables is zero.

To speed up the search procedure, we can use the rule \( \tilde{V} = V \) called the full exchange rule. This rule states that all variables F and G are exchanged that do not meet the Eq. (4.6) - (4.9), and the computational acceleration rule by reducing the number of iterations required until termination. However, such a rule can result cycles and fail to obtain optimal values, although these events are rare. To ensure the termination of the following exchange set for the Press (4.13), (4.14):

\[ \tilde{V} = \{ i : i = \text{maks} \{ i \in V \} \} \]

(4.15)

Supporting rules, where only improper variables with the largest indexes are exchanged, there are the main single-pivot rule. This rule guarantees finite termination. By incorporating full exchange rules or support roles, the main pivot method block for NNLS issues with one segment is obtained. The algorithm is listed in Algorithm 1 as follow:
4.1 Algorithm 1

Input : C ∈ Rp×q and b ∈ Rp
Output : x (∈ Rp×1) = arg min x ≥ 0 ||Cx - b||²

1. Initialize F = ∅, G = {1, ..., q}, x = 0, y = -Cᵀb, α = 3, β = q + 1
2. Calculate X_F and Y_G by Eq (4.10) and (4.11)
3. While (X_F, Y_G) is not worth doing
   4. Calculate V by Eq. (4.12)
   5. If |V| < β, Create β = |V|, α = 3 and V = V
   6. If |V| ≥ β and α ≥ 1, make α = α - 1 and V = V
   7. If |V| ≥ β and α = 0, make V by Eq. (5.15)
   8. Update F and G by Eq. (5.14), (5.14)
   9. Update X_F and Y_G by Eq. (5.10) and (5.11)
10. End while

In the case support rules slower than full exchange rules, it is only in use if the full exchange rules do not work. Final Termination Algorithm 1 is achieved by controlling the number of undeserved variables. The variable α is used as a buffer against the full exchange rules that are tried. If the full exchange rules increase the number of improper variables, α minus one.

After α becomes zero, the supporting rule is used so that it makes the number of variables eligible less than the smallest value reached, filled in β. This occurs in a finite number of steps because the buffer rule has finite termination properties. As soon as the buffer rule reaches the latest low number of undeserved variables, return to the full exchange rule. There are used as default values for α, which means that the full exchange rule is tied up to three times until it has an undeserved number of variables. Since the variable is not feasible systematically reduced, the algorithm stops in the number of finite steps.

Suppose, if we want to solve the following NNLS issues:

\[
\min_{x \geq 0} \|Cx - B\|_F^2 \tag{4.16}
\]

Where, C ∈ Rp×q, B ∈ Rp×x and X ∈ Rp×x. It can be done using Algorithm 1 for each vector of the right-hand side b₁, ..., b_r with B = [b₁, ..., b_r], since the X columns are not dependent on each other. But the computing is inefficient.

In Algorithm 1, the main computational load is more than the need to define X_F and Y_G as written in the Eq. (4.10) - (4.11). The Eq (4.10) can be solved by a normal equation:

\[C_F^T C_F X_F = C_F^T b \tag{4.17}\]

And Eq (4.11) can be written as:

\[Y_G = C_F^T C_F X_F - C_G^T b \tag{4.18}\]

Required C_F^T C_F, C_F^T b, C_F^T C_F, and C_G^T b to complete the Eq. (4.17) and (4.18) and these matrices and vectors vary throughout the iteration because F and G are cleared in the three iterations.

Note that for NNLS problems arising from the NMF, C matrix is very long and thin and calculates C_F^T C_F, C_F^T b, C_F^T C_F, and C_G^T b is computationally expensive. However, C^T C and C^T B can be started...
and reused later in the iteration. Since $C^T_F C_F, C^T_I C_I, G^T_F C_F,$ and $G^T_I C_I$ can be taken so that the submatrix of $C^T C$ and $C^T G$ is also small. This idea is used for the case of one right-hand side, but its influence is more observable in case many right-sideds.

Other developments arise from the fact that $X$ is flat and wide in the NNLS issue for NMF. Assuming, Algorithm 1 is run for vector with many right sides. In each iteration, there is an index set of $F_j$ and $G_j$ for each of the columns $j \in \{1, ..., r\}$. $X_p$, and $Y_0$ since they are computed using Eq. (4.17) and (4.18). The idea is to obtain a group of columns that share the same set of indexes $F_j$ and $G_j$ and complete the Eq. (4.17) to the columns in the same group. If done so, recurrent computation of Cholesky factorization is avoided upon completion of the Eq. (4.17). So, if $X$ is flat and wide, which is the case for the NNLS problem in NMF, more columns can share the set of $F_j$ and $G_j$ indexes, resulting in acceleration of the set process. The improved mainstream pivot method for many right-hand sections is listed in Algorithm 2.

4.2. Algorithm 2.

**Input**: $C \in \mathbb{R}^{n \times q}, B \in \mathbb{R}^{m \times r}$

**output**: $x \in \mathbb{R}^{n \times r}$ = \text{arg} \, \min_{x \geq 0} \|Cx - B\|^2_F

1. Calculate $C^T C$ and $C^T B$
2. Start $F_j = 0$ and $G_j = \{1, ..., q\} \forall j \in \{1, ..., r\}$.
   Create $X > 0, Y = -CTB, \alpha (\in \mathbb{R}^r) = 3$ and $\beta (\in \mathbb{R}^r) = q + 1$
3. Calculate $X_p$ and $Y_{G_j}$; For all $j \in \{1, ..., r\}$ by the Eq. (4.10) - (4.11) using column grouping.
4. While every $(X_p, Y_{G_j})$ is not worth doing
5. Specify a column index in the name of an ineligible settlement: $I = \{j: (X_p, Y_{G_j})$ improper\}
6. Calculate $\forall j \in I$ by the Eq. (4.12).
7. For all $j \in I$ with $|V_j| < B_I$, area $\beta j = |V_j|, A j = 3$ times $\hat{\beta} = \sum |V_j|$
8. For all $j \in I$ with $|V_j| \geq \beta j$ and $\alpha j \geq \Delta$, create, $\alpha j = \alpha j - 1$ and $\forall j = V_j$
9. For all $j \in I$ with $|V_j| \geq \beta j$ and $\alpha j = 0$, make $\bar{\beta}$ by the Eq. (4.15)
10. Update $F_j$ and $G_j$, $\forall j \in I$ with the Eq. (4.14)
11. Update $X_p$ and $Y_{C_j}, \forall j \in I$ with the Eq. (4.10), (4.11) which using column grouping
12. End while

In NMF, discontinuation criteria can be reviewed from whether the point of stasiment, which is done with the following criteria:

In accordance with the terms of the KKT, $(W, H)$ obtains the stasiment point of the Equation (4.2) if and only if

$$W \geq 0, H \geq 0 \quad \text{(4.19)}$$
$$\Delta f_W = 0, \Delta f_H = 0 \quad \text{(4.20)}$$
$$W * \nabla f_W = 0, \quad H * \nabla f_H = 0 \quad \text{(4.21)}$$

By * denoting multiplication of each element, define projected stasiment $\nabla^p f_W$ as $$(\nabla^p f_W) = \begin{cases} (\nabla f_W) & \text{if } (\nabla f_W) < 0 \text{ or } W_I > 0 \text{ if not and } \nabla^p f_W \text{ by analog, condition in Eq. (4.14)} - (4.21) \end{cases}$$

can be written again as $\nabla^p f_W = 0$ and $\nabla^p f_H = 0$. Use the projected gradient number defined as:

$$\Delta = \sqrt{||\nabla^p f_W||_F^2 + ||\nabla^p f_H||_F^2} \quad \text{(4.22)}$$
By using this definition, the termination criterion is

\[ \frac{\Delta}{\Delta_0} \leq \varepsilon \]  

(4.23)

Where \( \Delta_0 \) is the value of by marking the initial value of \( W \) and \( H \), \( \varepsilon \) the tolerance value to be selected.

5. Conclusion and Future Work

This study proposes a new algorithm to complete FMN based on the ANLS framework as an alternative algorithm model for data reduction. Such an algorithm is built on the upper-pivot method for the nonnegative least square problem. Methods include the exchange of several variables between sets in the process with the aim of reaching the final partition into the active and passive set quicker. The method is improved so that the power applies the reduced gradient method to handle multiple right-hand case cases of nonnegative least square issues efficiently. This new algorithm has the same convergence properties as ANLS framework and can be extended to FMN formulation constrained such that FMN matrix power sparse.

Further research that can be done is to add Quasi-Newton and Variable-Matric methods in addition to the reduced gradient method. The addition of this method will be able to further accelerate convergence.

6. REFERENCES

[1] B. Corona, M. Nakano, H. Pérez, "Adaptive Watermarking Algorithm for Binary Image Watermarks", Lecture Notes in Computer Science, Springer, pp. 207-215, 2004.
[2] A. Reddy and B. N. Chatterji, "A new wavelet based logo-watermarking scheme," Pattern Recognition Letters, vol. 26, pp. 1019-1027, 2005.
[3] F. Gonzalez and J. Hernandez, "A tutorial on Digital Watermarking", In IEEE annual Carnahan conference on security technology, Spain, 1999.
[4] D. Kunder, "Multi-resolution Digital Watermarking Algorithms and Implications for Multimedia Signals", Ph.D. thesis, university of Toronto, Canada, 2001.
[5] J. Eggers, J. Su and B. Girod, "Robustness of a Blind Image Watermarking Scheme", Proc. IEEE Int. Conf. on Image Proc., Vancouver, 2000.
[6] Barni M., Bartolini F., Piva A., Multichannel watermarking of color images, IEEE Transaction on Circuits and Systems of Video Technology 12(3) (2002) 142-156.
[7] Kundur D., Hatzinakos D., Towards robust logo watermarking using multiresolution image fusion, IEEE Transcations on Multimedia 6 (2004) 185-197.
[8] C.S. Lu, H.Y.M Liao, “Multipurpose watermarking for image authentication and protection,” IEEE Transaction on Image Processing, vol. 10, pp. 1579-1592, Oct. 2001.
[9] L. Ghouti, A. Bouridane, M.K. Ibrahim, and S. Boussakta, “Digital image watermarking using balanced multiwavelets”, IEEE Trans. Signal Process., 2006, Vol. 54, No. 4, pp. 1519-1536.
[10] P. Tay and J. Havlicek, "Image Watermarking Using Wavelets", in Proceedings of the 2002 IEEE, pp. II.258 – II.261, 2002.
[11] P. Kumswat, Ki. Attakitmongcol and A. Striaew, "A New Approach for Optimization in Image Watermarking by Using Genetic Algorithms", IEEE Transactions on Signal Processing, Vol. 53, No. 12, pp. 4707-4719, December, 2005.
[12] H. Daren, L. Jifuen, H. Jiwu, and L. Hongmei, "A DWT-Based Image Watermarking Algorithm", in Proceedings of the IEEE International Conference on Multimedia and Expo, pp. 429-432, 2001.
[13] C. Hsu and J. Wu, "Multi-resolution Watermarking for Digital Images", *IEEE Transactions on Circuits and Systems- II*, Vol. 45, No. 8, pp. 1097-1101, August 1998.

[14] R. Mehul, "Discrete Wavelet Transform Based Multiple Watermarking Scheme", in *Proceedings of the 2003 IEEE TENCON*, pp. 935-938, 2003.

[15] Tu, J.H., Rowley, C.W., Luchtenburg, D.M., Brunton, S.L., Kutz, J.N.: “On dynamic mode decomposition: Theory and applications”. Journal of Computational Dynamics, 2014. 1(2), 391–421)