IEICE Proceeding Series

A Modified Multiplicative Update Algorithm for Convex Quadratic Programming Problems with Nonnegativity Constraints

Jiro Katayama, Norikazu Takahashi

Vol. 1 pp. 431-434
Publication Date: 2014/03/17
Online ISSN: 2188-5079

2012 International Symposium on Nonlinear Theory and its Applications
NOLTA2012, Palma, Majorca, Spain, October 22-26, 2012

A Modified Multiplicative Update Algorithm for Convex Quadratic Programming Problems with Nonnegativity Constraints

Jiro Katayama† and Norikazu Takahashi‡

†Graduate School of Information Science and Electrical Engineering, Kyushu University
‡Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan.
Email: jiro.katayama@inf.kyushu-u.ac.jp, norikazu@inf.kyushu-u.ac.jp

Abstract—A multiplicative update for solving convex quadratic programming problems with nonnegativity constraints, which was proposed by Sha *et al.*, has three advantages: 1) nonnegativity of solutions is automatically satisfied, 2) no parameter tuning is needed, and 3) implementation is easy because of simple update formula. However, the global convergence of the update is not always guaranteed. In this paper, we propose a modified version of the multiplicative update and prove its global convergence without any assumption on the problem. We also show experimentally that our modification affects neither the computation time nor the number of iterations

1. Introduction

Problems of minimizing an objective function under nonnegativity constraints arise in various fields. For example, Nonnegative Matrix Factorization (NMF) [1, 2], which is to approximate a given large nonnegative matrix by the product of two small nonnegative matrices and has attracted remarkable attention in the fields of machine learning, signal processing and so on, is formulated as an optimization problem with nonnegativity constraints. In this paper, as an important class of these problems, we consider convex quadratic programming (QP) problems with nonnegativity constraints.

Sha *et al.* [3] recently proposed a multiplicative update for convex QP problems with nonnegativity constraints, which is based on multiplicative updates for NMF developed by Lee and Seung [1]. The algorithm of Sha *et al.* has three main advantages: 1) nonnegativity of variables is automatically satisfied, 2) no parameter tuning is needed, and 3) implementation is easy because of simple update formula. Furthermore, under some assumptions on the QP problem and the initial solution, Sha *et al.* proved that any sequence of solutions generated by their update converges to the unique optimal solution.

In this paper, we propose a modified version of the multiplicative update of Sha *et al.* and prove its global convergence without any assumption on the QP problem and the initial solution. We also construct an multiplicative update algorithm which always stops within a finite number of iterations after finding an approximate solution. We finally show experimentally that our modification affects neither the computation time nor the number of iterations.

2. Multiplicative Update Proposed by Sha et al.

We consider optimization problems of the form:

$$\begin{array}{ll} \text{minimize} & F(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} + \mathbf{b}^{\mathrm{T}}\mathbf{v} \\ \text{subject to} & \mathbf{v} \geq \mathbf{0} \end{array} \tag{1}$$

where $\mathbf{v} = [v_1, v_2, \ldots, v_n]^{\mathrm{T}} \in \mathbb{R}^n$ is a variable vector, $\mathbf{A} = [A_{ij}] \in \mathbb{R}^{n\times n}$ is a positive definite constant matrix, and $\mathbf{b} = [b_1, b_2, \ldots, b_n]^{\mathrm{T}} \in \mathbb{R}^n$ is a constant vector. Throughout this paper, the inequality between two vectors means componentwise inequality. Based on the idea behind the development of multiplicative updates for NMF [1], Sha *et al.* [3] recently proposed a multiplicative update for solving the optimization problem (1). Let $\mathbf{v}^k = [v_1^k, v_2^k, \ldots, v_n^k]^{\mathrm{T}}$ be the solution after k iterations. Then their update is expressed as follows:

$$v_i^{k+1} = \frac{-b_i + \sqrt{b_i^2 + 4a_i^k c_i^k}}{2a_i^k} v_i^k \tag{2}$$

where $a_i^k = (\mathbf{A}^+\mathbf{v}^k)_i$, $c_i^k = (\mathbf{A}^-\mathbf{v}^k)_i$, $\mathbf{A}^+ = [A_{ij}^+]$ is defined by

$$A_{ij}^+ = \begin{cases} A_{ij}, & \text{if } A_{ij} > 0 \\ 0, & \text{otherwise} \end{cases}$$

and $\mathbf{A}^+ = [A_{ij}^-]$ is defined by

$$A_{ij}^- = \begin{cases} |A_{ij}|, & \text{if } A_{ij} < 0 \\ 0, & \text{otherwise} \end{cases}$$

It is easily seen that the right-hand side of (2) is positive if $v_i^k > 0$ and the following assumption is satisfied.

Assumption 1 *The i-th row of* $\mathbf{A}$ *has one or more negative elements whenever* $b_i \geq 0$.

Therefore, if the initial solution $\mathbf{v}^0$ is positive and Assumption 1 is valid then any sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ generated by (2) satisfies $\mathbf{v}^k > \mathbf{0}$ for all k.

Sha *et al.* studied in detail the convergence properties of (2) under Assumption 1 and proved the following theorem by using Zangwill's global convergence theorem [6, 7].

Theorem 1 ([3]) Suppose that the origin is not the optimal solution of (1) and the initial vector $\mathbf{v}^0$ is positive and $F(\mathbf{v}^0) < F(\mathbf{0}) = 0$. Then the sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ converges to the optimal solution of (1).

Even if Assumption 1 is satisfied and the initial solution $\mathbf{v}^0$ is positive, the limit of the sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ may not be a positive vector. Therefore, the update must be defined for all nonnegative vectors $\mathbf{v}^k$ in order to apply Zangwill's global convergence theorem. However, in the update (2), it is not clear how the case where $a_i^k = 0$ is dealt with. As a possible solution, we consider the modified update:

$$v_i^{k+1} = \begin{cases} \frac{-b_i + \sqrt{b_i^2 + 4a_i^k c_i^k}}{2a_i^k} v_i^k, & \text{if } v_i^k > 0 \\ 0, & \text{if } v_i^k = 0 \end{cases} \tag{3}$$

Let us suppose that off-diagonal elements in the i-th row and i-th column of $\mathbf{A}$ are all zero. Then A_{ii} must be positive because $\mathbf{A}$ is assumed to be positive definite. Let us further suppose that b_i is negative. In this case, we have

$$\begin{aligned} \frac{-b_i + \sqrt{b_i^2 + 4a_i^k c_i^k}}{2a_i^k} v_i^k &= \frac{-b_i + \sqrt{b_i^2 + 0}}{2A_{ii} v_i^k} v_i^k \\ &= \frac{-b_i}{A_{ii}} > 0 \end{aligned}$$

for all $v_i^k > 0$ and hence the update (3) can be rewritten as

$$v_i^{k+1} = \begin{cases} -b_i/A_{ii}, & \text{if } v_i^k > 0 \\ 0, & \text{if } v_i^k = 0 \end{cases}$$

This means that the update (3) is not continuous in the region $\{\mathbf{v} \in \mathbb{R}^n \mid \mathbf{v} \geq \mathbf{0}, v_i = 0\}$, and hence Zangwill's global convergence theorem cannot be applied to (3).

3. Modified Multiplicative Update and Its Global Convergence

In order to avoid the problem described in the previous section, we propose a modified multiplicative update:

$$v_i^{k+1} = \max\left(\epsilon, \frac{-b_i + \sqrt{b_i^2 + 4a_i^k c_i^k}}{2a_i^k} v_i^k\right) \tag{4}$$

where ϵ is any positive constant. Note that the same idea is used in Reference [4, 5] in which a modified multiplicative update for NMF is proposed. It is apparent that if $\mathbf{v}^k > \mathbf{0}$ then the right-hand side is well-defined and we have $\mathbf{v}^{k+1} \geq \epsilon\mathbf{1}$ where $\mathbf{1} \in \mathbb{R}^n$ is the vector whose elements are all one. Therefore, if we choose the initial solution $\mathbf{v}^0$ such that $\mathbf{v}^0 \geq \epsilon\mathbf{1}$ then the sequence of solutions $\{\mathbf{v}^k\}_{k=0}^{\infty}$ generated by (4) satisfies $\mathbf{v}^k \geq \epsilon\mathbf{1}$ for all k. We thus consider in the following the optimization problem

$$\begin{array}{ll} \text{minimize} & F(\mathbf{v}) = \frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{v} + \mathbf{b}^{\mathrm{T}}\mathbf{v} \\ \text{subject to} & \mathbf{v} \geq \epsilon\mathbf{1} \end{array} \tag{5}$$

instead of (1). Note that (5) has the unique optimal solution for any positive constant ϵ and it approaches the optimal solution of (1) as ϵ goes to zero. As is well known in optimization theory, $\mathbf{v} \in [\epsilon, \infty)^n$ is the optimal solution of (5) if and only if the KKT conditions

$$\mathbf{A}\mathbf{v} + \mathbf{b} \geq \mathbf{0} \tag{6}$$

$$(\mathbf{A}\mathbf{v} + \mathbf{b})_i(-v_i + \epsilon) = 0, \quad i = 1, 2, \ldots, n \tag{7}$$

are satisfied.

One of the main results of this paper is given by the following theorem.

Theorem 2 For any positive constant ϵ and the initial solution $\mathbf{v}^0 \in [\epsilon, \infty)^n$, the sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ generated by the modified multiplicative update (4) has at least one convergent subsequence and they converge to the unique optimal solution of (5).

Note that this theorem does not require any assumption on $\mathbf{A}$ and $\mathbf{b}$ except that $\mathbf{A}$ is positive definite. Proof of this theorem will be given in the next section.

We next develop a multiplicative update algorithm which always stops within a finite number of iterations. The stopping criterion we employ for the algorithm is given by

$$\mathbf{A}\mathbf{v} + \mathbf{b} \geq -\delta_1\mathbf{1} \tag{8}$$

$$|(\mathbf{A}\mathbf{v} + \mathbf{b})_i(-v_i + \epsilon)| < \delta_2, \quad i = 1, 2, \ldots, n \tag{9}$$

where δ_1 and δ_2 are positive constants. As is easily seen, this criterion is a relaxed version of the KKT conditions (6) and (7). The proposed algorithm is described as follows.

Algorithm 1

Step 1: Set $k = 0$ and choose the initial solution $\mathbf{v}^0$ such that $\mathbf{v}^0 \geq \epsilon\mathbf{1}$.

Step 2: Find $\mathbf{v}^{k+1}$ by the modified update (4).

Step 3: If $\mathbf{v} = \mathbf{v}^{k+1}$ satisfies the stopping criterion (8) and (9) then stop. Otherwise, add 1 to k and go to Step 2.

From Theorem 1 and the continuity of the stopping criterion, we have the following theorem.

Theorem 3 For any positive constants ϵ, δ_1 and δ_2, Algorithm 1 stops within a finite number of iterations.

Proof: Let $\{\mathbf{v}^{k_l}\}_{l=0}^{\infty}$ be any convergent subsequence of the sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ generated by the update (4). By Theorem 1, it converges to the unique optimal solution of (5). We therefore have

$$\lim_{l\to\infty}\left(\mathbf{A}\mathbf{v}^{k_l} + \mathbf{b}\right) \geq \mathbf{0}$$

$$\lim_{l\to\infty}\left\{(\mathbf{A}\mathbf{v}^{k_l} + \mathbf{b})_i(-v_i^{k_l} + \epsilon)\right\} = 0, \quad i = 1, 2, \ldots, n$$

In other words, for any positive constants δ_1 and δ_2, there exists an positive integer L such that

$$\mathbf{A}\mathbf{v}^{k_l} + \mathbf{b} \geq -\delta_1\mathbf{1}$$

$$|(\mathbf{A}\mathbf{v}^{k_l} + \mathbf{b})_i(-v_i^{k_l} + \epsilon)| \leq \delta_2, \quad i = 1, 2, \ldots, n$$

for all $l \geq L$. This completes the proof. □

4. Proof of Theorem 1

We will prove Theorem 1 by using the same approach as Sha *et al.* [3], that is, we apply Zangwill's global convergence theorem [6, 7] to the update (4). We hereafter express (4) as $\mathbf{v}^{k+1} = \mathcal{M}(\mathbf{v}^k)$ for notational simplicity, where $\mathcal{M}$ is a mapping from $[\epsilon, \infty)^n$ into itself. In order to prove Theorem 2 by using Zangwill's global convergence theorem, it suffices for us to show that the mapping $\mathcal{M}$ satisfies the following conditions.

(i) For all $k \geq 0$, $\mathbf{v}^k$ belongs to a compact set.

(ii) The mapping $\mathcal{M}$ satisfies

  (a) $\mathbf{v} \neq \mathbf{v}^* \Rightarrow F(\mathcal{M}(\mathbf{v})) < F(\mathbf{v})$

  (b) $\mathbf{v} = \mathbf{v}^* \Rightarrow F(\mathcal{M}(\mathbf{v})) \leq F(\mathbf{v})$

(iii) The mapping $\mathcal{M}$ is continuous for all vectors $\mathbf{v} \in [\epsilon, \infty)^n$ except the unique optimal solution of (5).

Lemma 1 The mapping $\mathcal{M}$ is continuous in $[\epsilon, \infty)^n$.

Lemma 2 If $\mathbf{v} \in [\epsilon, \infty)^n$ is not a fixed point of the mapping $\mathcal{M}$ then $F(\mathcal{M}(\mathbf{v})) < F(\mathbf{v})$.

Proof: For given $\mathbf{v} \in [\epsilon, \infty)^n$, we consider the optimization problem

$$\begin{array}{ll} \text{minimize} & G(\mathbf{u}, \mathbf{v}) \\ \text{subject to} & \mathbf{u} \geq \epsilon \mathbf{1} \end{array} \tag{10}$$

where

$$\begin{aligned} G(\mathbf{u}, \mathbf{v}) \quad = \quad & \frac{1}{2} \sum_{i=1}^{n} \frac{(\mathbf{A}^+\mathbf{v})_i}{v_i} u_i^2 \\ & - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij}^- v_i v_j \left(1 + \log \frac{u_i u_j}{v_i v_j}\right) + \sum_{i=1}^{n} b_i u_i \end{aligned}$$

which is well defined in $[\epsilon, \infty)^n$ and continuous everywhere in $[\epsilon, \infty)^n$. The function $G(\mathbf{u}, \mathbf{v})$ is an auxiliary function of $F(\mathbf{v})$, that is, it satisfies

$$F(\mathbf{u}) \leq G(\mathbf{u}, \mathbf{v}), \quad \forall \mathbf{u}, \mathbf{v} \geq \epsilon \mathbf{1} \tag{11}$$

$$F(\mathbf{v}) = G(\mathbf{v}, \mathbf{v}), \quad \forall \mathbf{v} \geq \epsilon \mathbf{1} \tag{12}$$

(see Reference [3] for details). Also, since $G(\mathbf{u}, \mathbf{v})$ can be rewritten as

$$G(\mathbf{u}, \mathbf{v}) = \sum_{i=1}^{n} G_i(u_i) - \frac{1}{2} \mathbf{v}^{\mathrm{T}} \mathbf{A}^- \mathbf{v}$$

where

$$G_i(u_i) = \frac{1}{2} \frac{(\mathbf{A}^+\mathbf{v})_i}{v_i} u_i^2 - (\mathbf{A}^-\mathbf{v})_i v_i \log \frac{u_i}{v_i} + b_i u_i$$

the optimization problem (10) is decomposed into n independent optimization problems of the form:

$$\begin{array}{ll} \text{minimize} & G_i(u_i) \\ \text{subject to} & u_i \geq \epsilon \end{array} \tag{13}$$

The objective function $G_i(u_i)$ is strictly convex in $[\epsilon, \infty)$ because

$$G_i''(u_i) = \frac{(\mathbf{A}^+\mathbf{v})_i}{v_i} + \frac{(\mathbf{A}^-\mathbf{v})_i v_i}{u_i^2} > 0\,.$$

Thus the optimal solution u_i^* of (13) is uniquely determined by considering the solution of the equation $G_i'(u_i) = 0$, and is given by

$$u_i^* = \max\left(\epsilon, \frac{-b_i + \sqrt{b_i^2 + 4(\mathbf{A}^+\mathbf{v})_i(\mathbf{A}^-\mathbf{v})_i}}{2(\mathbf{A}^+\mathbf{v})_i} v_i\right).$$

This implies that the optimal solution of (10) is given by $\mathcal{M}(\mathbf{v})$. Therefore, we have $G(\mathcal{M}(\mathbf{v}), \mathbf{v}) < G(\mathbf{v}, \mathbf{v})$ if $\mathcal{M}(\mathbf{v}) \neq \mathbf{v}$. From this inequality and the properties (11) and (12), we have

$$F(\mathcal{M}(\mathbf{v})) \leq G(\mathcal{M}(\mathbf{v}), \mathbf{v}) < G(\mathbf{v}, \mathbf{v}) = F(\mathbf{v})$$

which completes the proof. □

Lemma 3 $\mathbf{v} \in [\epsilon, \infty)^n$ is a fixed point of the mapping $\mathcal{M}$ if and only if it is the unique optimal solution of (5).

Proof: A vector $\mathbf{v} \in [\epsilon, \infty)^n$ is a fixed point of $\mathcal{M}$ if and only if

$$\frac{-b_i + \sqrt{b_i^2 + 4(\mathbf{A}^+\mathbf{v})_i(\mathbf{A}^-\mathbf{v})_i}}{2(\mathbf{A}^+\mathbf{v})_i} v_i \begin{cases} \leq v_i, & \text{if } v_i = \epsilon \\ = v_i, & \text{if } v_i > \epsilon \end{cases}, \quad i = 1, 2, \ldots, n$$

which can be rewritten as

$$(\mathbf{A}^+\mathbf{v})_i + b_i - (\mathbf{A}^-\mathbf{v})_i \begin{cases} \geq 0, & \text{if } v_i = \epsilon \\ = 0, & \text{if } v_i > \epsilon \end{cases}, \quad i = 1, 2, \ldots, n$$

Noting that $(\mathbf{A}^+\mathbf{v})_i + b_i - (\mathbf{A}^-\mathbf{v})_i = (\mathbf{A}\mathbf{v} + \mathbf{b})_i$, we see that this is equivalent to (6) and (7). Since $\mathbf{v} \in [\epsilon, \infty)^n$ is the optimal solution of (5) if and only if (6) and (7) are satisfied, we can conclude that $\mathbf{v} \in [\epsilon, \infty)^n$ is a fixed point of $\mathcal{M}$ if and only if $\mathbf{v}$ is the unique optimal solution of (5). □

From Lemmas 2 and 3, we have the following lemma.

Lemma 4 If $\mathbf{v} \in [\epsilon, \infty)^n$ is the optimal solution of (5) then $F(\mathcal{M}(\mathbf{v})) = F(\mathbf{v})$ holds because $\mathbf{v}$ is the fixed point of $\mathcal{M}$. Otherwise, $F(\mathcal{M}(\mathbf{v})) < F(\mathbf{v})$ holds.

Lemma 5 For any initial solution $\mathbf{v}^0 \in [\epsilon, \infty)^n$, the sequence $\{\mathbf{v}^k\}_{k=0}^{\infty}$ generated by (4) belongs to the compact set $S = \{\mathbf{v} \in \mathbb{R}^n \,|\, \mathbf{v} \geq \epsilon \mathbf{1}, F(\mathbf{v}) \leq F(\mathbf{v}^0)\}$.

Proof: It follows from Lemma 4 that $F(\mathbf{v}^k) \leq F(\mathbf{v}^0)$ holds for all k, that is, $\mathbf{v}^k \in S$ for all k. Since $F(\mathbf{v})$ is continuous and strongly convex, the level set $\{\mathbf{v} \in \mathbb{R}^n \,|\, F(\mathbf{v}) \leq F(\mathbf{v}^0)\}$ is compact. Therefore S is also compact. □

Lemmas 1, 4 and 5 mean that all of the three conditions in Zangwill's global convergence theorem are satisfied. This completes the proof of Theorem 1.

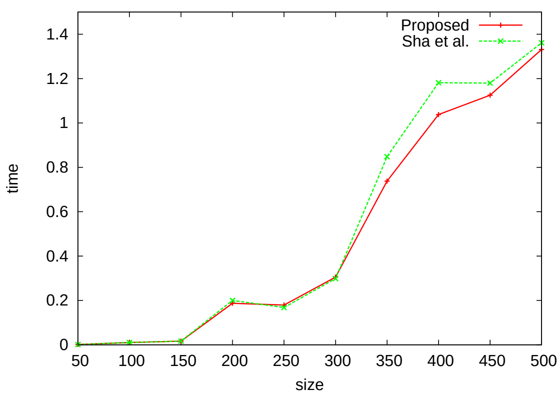

Figure 1: Comparison of computation time between the original (green) and modified (red) updates.

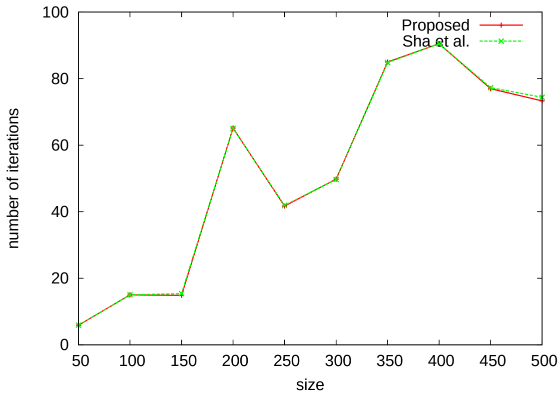

Figure 2: Comparison of the number of iterations between the original (green) and modified (red) updates.

5. Numerical Experiments

In order to examine the effect of the modifications of the multiplicative update and the optimization problem on the computational cost, we conducted numerical experiments. We implemented both the original update (2) and the modified one (4) with Scilab 5.3.3 and applied them to convex QP problems with nonnegativity constraints with size n from 50 to 500 on a Windows PC with Intel Core i5 2.53GHz CPU and 2GB RAM. The effect is evaluated by the computation time and the number of iterations which are obtained by computing the averages over 30 randomly generated problems. In all experiments, the positive constant ϵ was set to 10^{-4}, and the positive constants δ_1 and δ_2 are set to 10^{-5}.

Figure 1 shows the computation time of the original and modified updates for QP problems with various sizes. Although the modified update requires an additional operation, i.e., the max operation, no significant difference can be seen from the results. Moreover, when n is between 350 and 450, the modified update is a little bit faster than the original. Unfortunately, the reason for this is not clear to the authors. But anyway, we can say that our modification has little effect on the computation time. Figure 2 shows the number of iterations of the original and modified updates. For all values of n, there is little difference between these two updates. We can say from the results that our modification has no effect on the number of iterations.

6. Conclusion

Global convergence of the multiplicative update proposed by Sha *et al.* [3] for solving strictly convex QP problems with nonnegativity constraints was studied. We have first pointed out an error in their global convergence proof. We have next proposed a modified version of the multiplicative update and proved its global convergence. We have finally shown that the modification has little effect on the computation time and the number of iterations. A future problem is to extend the proposed update to the case where the objective function is not strictly convex.

Acknowledgments

This work was supported by KAKENHI 24560076 and 23310104.

References

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in T. K. Leen, T. G. Dietterich and V. Tresp (Eds.), *Advances in Neural Information Processing Systems*, vol. 13, pp. 556–562, 2001.

[2] A. Cichocki, R. Zdunek, A. H. Phan and S.-I. Amari, *Nonnegative Matrix and Tensor Factorizations*, John Wiley & Sons, 2009.

[3] F. Sha, Y. Lin, L. K. Saul, D. D. Lee, "Multiplicative updates for nonnegative quadratic programming," *Neural Computation*, vol. 19, pp. 2004–2031, 2007.

[4] N. Gillis and F. Glineur, "Nonnegative factorization and the maximum edge biclique problem," *ArXiv*, 0810.4225, Oct. 2008.

[5] R. Hibi and N. Takahashi, "A modified multiplicative update algorithm for Euclidean distance-based nonnegative matrix factorization and its global convergence," *Proc. 18th International Conference on Neural Information Processing*, Part-II, pp. 655–662, 2011.

[6] W. I. Zangwill, *Nonlinear Programming: A Unified Approach*, Prentice-Hall, 1969.

[7] D. G. Luenberger, *Linear and Nonlinear Programming*, Second Edition, Addison-Wesley Publishing Company, 1984.