

# Nonnegative Orthogonal Graph Matching

Bo Jiang,<sup>1</sup> Jin Tang,<sup>1</sup> Chris Ding,<sup>2,1</sup> Bin Luo<sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, Anhui University, Hefei, 230601, China

<sup>2</sup>CSE Department, University of Texas at Arlington, Arlington, TX 76019, USA

jiangbo@ahu.edu.cn, ahhftang@gmail.com, chqding@uta.edu, luobin@ahu.edu.cn

## Abstract

Graph matching problem that incorporates pair-wise constraints can be formulated as Quadratic Assignment Problem (QAP). The optimal solution of QAP is discrete and combinatorial, which makes QAP problem NP-hard. Thus, many algorithms have been proposed to find approximate solutions. In this paper, we propose a new algorithm, called Nonnegative Orthogonal Graph Matching (NOGM), for QAP matching problem. NOGM is motivated by our new observation that the discrete mapping constraint of QAP can be equivalently encoded by a *nonnegative orthogonal* constraint which is much easier to implement computationally. Based on this observation, we develop an effective multiplicative update algorithm to solve NOGM and thus can find an effective approximate solution for QAP problem. Comparing with many traditional continuous methods which usually obtain continuous solutions and should be further discretized, NOGM can obtain a sparse solution and thus incorporates the desirable discrete constraint naturally in its optimization. Promising experimental results demonstrate benefits of NOGM algorithm.

## Introduction

Many problems of interest in computer vision and machine learning area can be formulated as a problem of finding consistent correspondences between two sets of features. This problem can usually be formulated and solved by graph matching model. From optimization aspect, previous approaches have formulated graph matching as a Quadratic Assignment Problem (QAP).

It is known that the optimal solution of QAP matching problem should satisfy both discrete and mapping (one-to-one or one-to-many) constraints simultaneously. This combinatorial constraints make QAP problem NP-hard. Thus many algorithms have been proposed to find approximate solutions for graph matching problem (Enqvist, Josephon, and Kahl 2009; Leordeanu, Hebert, and Sukthankar 2009; Zhou and la Torre 2012; van Wyk and van Wyk 2004; Zaslavskiy, Bach, and Vert 2009; Liu, Qiao, and Xu 2012; Zhang et al. 2016; Adamczewski, Suh, and Lee 2015). Among them, one kind of popular methods is to use continuous optimization techniques (Gold and A 1996; Leordeanu

and Hebert 2005; Cour, Srinivasan, and Shi 2006; Enqvist, Josephon, and Kahl 2009; Choi and Kweon 2009; Cho, Lee, and Lee 2010). These methods generally first develop an approximate continuous problem for QAP by relaxing its discrete mapping constraint and then find the optimal solution for this continuous problem. At last, they use some discretization techniques to obtain the final permutation solution (Cho, Lee, and Lee 2010; Leordeanu and Hebert 2005). One drawback of these continuous algorithms is that the discrete constraint of QAP matching problem is entirely ignored and thus need to be further discretized to obtain a discrete solution for the problem. Obviously, this post-discretization step is independent of QAP matching objective and thus leads to weak local optimal solution (Jiang et al. 2017; Zhou and la Torre 2012). In addition to continuous algorithm, Leordeanu and Hebert (Leordeanu, Hebert, and Sukthankar 2009) also propose an iterative matching method (IPFP) which optimizes the QAP in the discrete domain and thus finds a discrete solution strictly in the optimization process. In this paper, we focus on continuous optimization methods.

This paper proposes a new algorithm, called Nonnegative Orthogonal Graph Matching (NOGM), for QAP problem. NOGM is based on our new observation that the discrete mapping constraint of QAP can be equivalently encoded by a nonnegative orthogonal constraint. This motivates us to develop a multiplicative update algorithm to find an effective solution for QAP. Comparing with many traditional continuous methods which usually obtain continuous solutions and need to be further discretized, NOGM can generate a sparse solution for the problem and thus better incorporates the desirable discrete constraint in its optimization. To the best of our knowledge, this nonnegative orthogonal constraint has not been studied or emphasized in QAP problem. Promising experimental results demonstrate the benefits and effectiveness of the proposed NOGM method.

## Problem Formulation and Related Works

Assume that two graphs to be matched are  $G^D = (V^D, E^D)$  and  $G^M = (V^M, E^M)$  where  $V$  represents a set of nodes and  $E$  denotes edges. The graph matching problem, in its most recent and general form, can be formulated as Quadratic Assignment Problem (QAP) problem, i.e., finding the indicator matrix  $\mathbf{X} \in \{0, 1\}^{n \times n}$  where

$n = |G^D| = |G^M|$  that maximizes the following quadratic score function (Leordeanu, Hebert, and Sukthankar 2009; Cho, Lee, and Lee 2010),

$$\max_{\mathbf{X}} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \mathbf{W}_{ij,kl} \mathbf{X}_{ij} \mathbf{X}_{kl} = \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

s.t.  $\mathbf{X}\mathbf{1} = \mathbf{1}, \mathbf{X}^T \mathbf{1} = \mathbf{1}, \mathbf{X}_{ij} \in \{0, 1\}.$  (1)

where  $\mathbf{W}_{ij,kl}$  measures how similar the pair of nodes  $(v_i, v_k) \in G^D$  is with the pair of their corresponding nodes  $(v_j, v_l) \in G^M$  in both local appearance and pair-wise geometry. The two-way affine constraints ensure the one-to-one matching between nodes of graph  $G^D$  and  $G^M$  (Cho, Lee, and Lee 2010; Gold and A 1996).

The above QAP problem (Eq.(1)) with the discrete constraint is NP-hard. Thus relaxation models and algorithms are required to find approximate solutions (Conte et al. 2004; Zhou and la Torre 2012; Leordeanu and Hebert 2005; Cour, Srinivasan, and Shi 2006). Here, we briefly review some most related relaxation works.

**Spectral Matching (SM):** Leordeanu et al. (Leordeanu and Hebert 2005) proposed to solve the following relaxation problem,

$$\max_{\mathbf{X}} \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

$$\text{s.t. } \text{vec}(\mathbf{X})^T \text{vec}(\mathbf{X}) = 1. \quad (2)$$

One main advantage of SM relaxation is that it has a global closed-form solution which is the leading eigenvector of  $\mathbf{W}$ . **Balanced Graph Matching:** Cour et al. (Cour, Srinivasan, and Shi 2006) presented a tighter relaxation by further incorporating the affine mapping constraint into SM relaxation (SMAC),

$$\max_{\mathbf{X}} \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

$$\text{s.t. } \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x}^T \mathbf{x} = 1, \quad (3)$$

where  $\mathbf{x} = \text{vec}(\mathbf{X})$ . Similar to SM, SMAC also has a closed-form eigenvector solution. However, the optimal solution is not necessarily nonnegative in SMAC.  $\mathbf{A}$  and  $\mathbf{b}$  are defined to make constraint  $\mathbf{A}\mathbf{x} = \mathbf{b}$  represent the doubly stochastic constraint.

**Re-weighted Random Walk Matching (RRWM):** Cho et al (Cho, Lee, and Lee 2010) introduced a probabilistic algorithm to interpret and find a local optimal solution for the following general relaxation model, i.e., solving the matching problem in the doubly (sub-)stochastic domain,

$$\max_{\mathbf{X}} \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

$$\text{s.t. } \mathbf{X}\mathbf{1} = \mathbf{1}, \mathbf{X}^T \mathbf{1} = \mathbf{1}, \mathbf{X}_{ij} \geq 0. \quad (4)$$

Many other algorithms, such as graduated assignment (GA) (Gold and A 1996), POCS (van Wyk and van Wyk 2004) etc, have also been developed to solve this problem.

<sup>1</sup>Here, we focus on equal size graph matching. For the graphs with different sizes, one can transform it to equal size problem by adding dummy nodes to the smaller graph.

The optimal solutions generated by the above methods are usually continuous and thus should be abruptly discretized to obtain the final discrete solutions. This discretization step is independent of original graph matching objective and thus usually leads to weak local optimum for the original problem. This is one main drawback of these methods.

## Nonnegative Orthogonal Graph Matching

In this section, we show that the discrete mapping constraint of Eq.(1) can be equivalently encoded by a nonnegative orthogonal constraint. This can be seen as follows.

Using orthogonal constraint, the above QAP problem (Eq.(1)) can be equivalently formulated as,

$$\max_{\mathbf{X}} \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

$$\text{s.t. } \mathbf{X}\mathbf{X}^T = \mathbf{I}, \mathbf{X}_{ij} \in \{0, 1\}.$$

Similar to constraint of Eq.(1), the orthogonal constraint here is used to encode the one-to-one mapping between two graphs. The discrete constraint in this problem is also combinatorial and thus difficult to implement computationally. One natural relaxation to this problem is to replace the discrete constraint  $\mathbf{X}_{ij} \in \{0, 1\}$  with nonnegative constraint and solve the following approximate problem, i.e.,

$$\max_{\mathbf{X}} \text{vec}(\mathbf{X})^T \mathbf{W} \text{vec}(\mathbf{X})$$

$$\text{s.t. } \mathbf{X}\mathbf{X}^T = \mathbf{I}, \mathbf{X}_{ij} \geq 0. \quad (5)$$

We call it as Nonnegative Orthogonal Graph Matching (NOGM). We show that NOGM is equivalent to Eq.(8).

**Proposition 1** For any matrix  $\mathbf{X}, \mathbf{X} \in \mathbb{R}^{n \times n}$ , if  $\mathbf{X}$  is non-negative orthogonal, i.e.,  $\mathbf{X}\mathbf{X}^T = \mathbf{I}, \mathbf{X}_{ij} \geq 0$ , then  $\mathbf{X}$  is a permutation matrix.

**Proof.** (1) Under the nonnegative constraint  $\mathbf{X}_{ij} \geq 0$ , the orthogonal constraint  $\mathbf{X}\mathbf{X}^T = \mathbf{I}$  indicates that each row  $\mathbf{X}_i$  of solution matrix  $\mathbf{X}$  has only one non-zero element. (2) The orthogonal constraint  $\mathbf{X}\mathbf{X}^T = \mathbf{I}$  also enforces  $\mathbf{X}_i \cdot \mathbf{X}_i^T = 1$ , which indicates that the only one non-zero element in each row  $\mathbf{X}_i$  should be equal to 1. Thus,  $\mathbf{X}_{ij} \in \{0, 1\}$ .  $\square$

**Remark.** Comparing with discrete constraint in original problem Eq.(8), the nonnegative constraint in NOGM (Eq.(9)) is much easier to implement computationally. To the best of our knowledge, this particular observation has not been explored before. Our aim in the following is to derive an effective multiplicative algorithm to solve NOGM problem approximately and thus can obtain an approximate solution for QAP problem.

## Optimization

Recently, multiplicative update algorithms have been widely adopted in many optimization problems with nonnegative constraint (Li et al. 2007; Lee and Seung 2001; Ding, Li, and Jordan 2010; 2008; Jiang et al. 2015; Ding et al. 2006; Chen, Wang, and Zhang 2007). Inspired by these works, in this section we propose develop a new multiplicative update algorithm to solve the proposed NOGM problem.

## Update algorithm

Starting from an initial matrix solution  $\mathbf{X}^{(0)}$ , the proposed algorithm iteratively updates a current solution  $\mathbf{X}^{(t)}$  as follows,

$$\mathbf{X}_{ij}^{(t+1)} = \mathbf{X}_{ij}^{(t)} \left( \frac{\mathbf{K}_{ij}^{(t)}}{(\Delta \mathbf{X}^{(t)})_{ij}} \right)^{1/2}, \quad (10)$$

where the multiplicatives  $\Delta$  is computed as

$$\Delta = \frac{1}{2}(\mathbf{K}^{(t)}\mathbf{X}^{(t)\top} + \mathbf{X}^{(t)}\mathbf{K}^{(t)\top}), \quad (11)$$

and  $\mathbf{K}^{(t)}$  is the matrix form of the vector  $(\mathbf{W} \text{vec}(\mathbf{X}^{(t)}))$ , i.e.,  $\mathbf{K}^{(t)} = \text{mat}(\mathbf{W} \text{vec}(\mathbf{X}^{(t)}))$ . The iteration starts with an initial  $\mathbf{X}^{(0)}$  and is repeated until convergence. Since both  $\mathbf{K}^{(t)}$  and  $\Delta$  are nonnegative (because  $\mathbf{W}$  is a real nonnegative symmetric matrix for the matching problem), thus the nonnegativity of  $\mathbf{X}^{(t)}$  is always guaranteed.

## Theoretical analysis

**Theorem 1** *Under the update rule of Eq.(10), the convergence solution satisfies the first-order Karush-Kuhn-Tucker (KKT) optimality condition.*

**Proof.** The Lagrangian function is

$$\mathcal{L}(\mathbf{X}) = \text{vec}(\mathbf{X})^\top \mathbf{W} \text{vec}(\mathbf{X}) - \text{Tr} \Delta(\mathbf{X}\mathbf{X}^\top - \mathbf{I}). \quad (12)$$

Note that, since  $\mathbf{X}\mathbf{X}^\top - \mathbf{I}$  is symmetric, thus  $\Delta$  is symmetric, i.e.,  $\Delta = \Delta^\top$ . Thus,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}_{ij}} = 2(\mathbf{K} - \Delta \mathbf{X})_{ij}. \quad (13)$$

where  $\mathbf{K}$  is the matrix form of the vector  $(\mathbf{W} \text{vec}(\mathbf{X}))$ , i.e.,  $\mathbf{K} = \text{mat}(\mathbf{W} \text{vec}(\mathbf{X}))$ . This leads to the following KKT complementary slackness condition,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}_{ij}} \mathbf{X}_{ij} = 2(\mathbf{K} - \Delta \mathbf{X})_{ij} \mathbf{X}_{ij} = 0. \quad (14)$$

Because  $\mathbf{X}\mathbf{X}^\top = \mathbf{I}$ , summing over index  $j$ , we can obtain the diagonal elements of the Lagrangian multipliers  $\Delta$  as,

$$(\mathbf{K}\mathbf{X}^\top)_{ii} = (\Delta \mathbf{X}\mathbf{X}^\top)_{ii} = \Delta_{ii} \quad (15)$$

In order to obtain the off-diagonal elements of  $\Delta$ , we ignore the nonnegative constraint of  $\mathbf{X}$  temporarily and set,

$$\frac{\partial \mathcal{L}(\mathbf{X})}{\partial \mathbf{X}_{ij}} = 0. \quad (16)$$

Thus, we have

$$\Delta_{ij} = (\mathbf{K}\mathbf{X}^\top)_{ij}, (i \neq j). \quad (17)$$

Integrating Eq.(15) and Eq.(17), we have a closed form solution for the Lagrangian multipliers  $\Delta$  as

$$\Delta = \mathbf{K}\mathbf{X}^\top. \quad (18)$$

Since  $\Delta$  is symmetric, thus we further set  $\Delta = \frac{1}{2}(\mathbf{K}\mathbf{X}^\top + \mathbf{X}\mathbf{K}^\top)$ . Clearly, the fixed points of the update rule Eq.(10)

satisfy  $(\mathbf{K} - \Delta \mathbf{X})_{ij} \mathbf{X}_{ij}^2 = 0$ , which is identical to Eq.(14).  $\square$

**Remark.** It is noted that in Eq.(18), the off-diagonal elements of the Lagrangian multipliers  $\Delta_{ij}, (i \neq j)$  are computed approximately. Therefore, the converged solution  $\mathbf{X}$  does not satisfy the orthogonal constraint  $(\mathbf{X}\mathbf{X}^\top = \mathbf{I})$  strictly. In other words, update algorithm Eq.(10) generally returns an approximate solution for the proposed NOGM problem (Eq.(9)). However, although approximately,  $\mathbf{X}$  can still satisfy the orthogonal constraint  $(\mathbf{X}\mathbf{X}^\top = \mathbf{I})$  strongly in the nonnegative domain and thus closes to a discrete solution. This is one important benefit of the proposed update algorithm and will be further illustrated in the following section in detail.

The convergence of our algorithm is guaranteed by the following theorem.

**Theorem 2** *Under the update rule of Eq.(10), the Lagrangian function  $\mathcal{L}(\mathbf{X})$  of Eq.(12) is monotonically increasing.*

**Proof.** We use the auxiliary function approach (Ding, Li, and Jordan 2010; Lee and Seung 2001). An auxiliary function  $Z(\mathbf{X}, \mathbf{X}')$  of Lagrangian function  $\mathcal{L}(\mathbf{X})$  satisfies following,

$$Z(\mathbf{X}, \mathbf{X}) = \mathcal{L}(\mathbf{X}), Z(\mathbf{X}, \mathbf{X}') \leq \mathcal{L}(\mathbf{X}). \quad (19)$$

We define

$$\mathbf{X}^{(t+1)} = \arg \max_{\mathbf{X}} Z(\mathbf{X}, \mathbf{X}^{(t)}). \quad (20)$$

Then by construction, we have

$$\mathcal{L}(\mathbf{X}^{(t)}) = Z(\mathbf{X}^{(t)}, \mathbf{X}^{(t)}) \leq Z(\mathbf{X}^{(t+1)}, \mathbf{X}^{(t)}) \leq \mathcal{L}(\mathbf{X}^{(t+1)}). \quad (21)$$

This proves that  $\mathcal{L}(\mathbf{X}^{(t)})$  is monotonically increasing.

In the following, we first give an appropriate auxiliary function and then find the global maximum of the auxiliary function. We rewrite Eq.(12) as

$$\mathcal{L}(\mathbf{X}) = \sum_{ij} \sum_{kl} \mathbf{W}_{ij,kl} \mathbf{X}_{ij} \mathbf{X}_{kl} - \text{Tr} \Delta(\mathbf{X}\mathbf{X}^\top - \mathbf{I}) \quad (22)$$

We can show that one auxiliary function of  $\mathcal{L}(\mathbf{X})$  can be defined as

$$Z(\mathbf{X}, \mathbf{X}') = \sum_{ij} \sum_{kl} \mathbf{W}_{ij,kl} \mathbf{X}'_{ij} \mathbf{X}'_{kl} (1 + \log \frac{\mathbf{X}_{ij} \mathbf{X}_{kl}}{\mathbf{X}'_{ij} \mathbf{X}'_{kl}}) - \sum_i \sum_j \frac{(\Delta \mathbf{X}')_{ij} \mathbf{X}_{ij}^2}{\mathbf{X}'_{ij}}. \quad (23)$$

Using the inequality  $z \geq 1 + \log z$  and setting  $Z = \frac{\mathbf{X}_{ij} \mathbf{X}_{kl}}{\mathbf{X}'_{ij} \mathbf{X}'_{kl}}$ , the first term of Eq.(23) is a lower bound of the first term in Eq.(22). Also, for any positive matrices  $\mathbf{A}, \mathbf{B}, \mathbf{S}$  and  $\mathbf{S}'$  with  $\mathbf{A}$  and  $\mathbf{B}$  symmetric, the following inequality always holds (Ding, Li, and Jordan 2010; 2008),

$$\sum_i \sum_j \frac{(\mathbf{A}\mathbf{S}'\mathbf{B})_{ij} \mathbf{S}_{ij}^2}{\mathbf{S}'_{ij}} \geq \text{Tr}(\mathbf{S}^\top \mathbf{A} \mathbf{S} \mathbf{B}) = \text{Tr}(\mathbf{S} \mathbf{B} \mathbf{S}^\top \mathbf{A}). \quad (24)$$

Using this equality, we can see that the second term in Eq.(23) is a lower bound of the second term in Eq.(22). Thus,  $Z(\mathbf{X}, \tilde{\mathbf{X}})$  is an auxiliary function of  $\mathcal{L}(\mathbf{X})$ .

According to Eq.(20), we need to find the global maximum of  $Z(\mathbf{X}, \mathbf{X}')$  for  $\mathbf{X}$ . The gradient is

$$\frac{\partial Z(\mathbf{X}, \mathbf{X}')}{\partial \mathbf{X}_{ij}} = 2 \left[ \frac{(\mathbf{W} \text{vec}(\mathbf{X}'))_{ij} \mathbf{X}'_{ij}}{\mathbf{X}_{ij}} - \frac{(\Delta \mathbf{X}')_{ij} \mathbf{X}_{ij}}{\mathbf{X}'_{ij}} \right] \quad (25)$$

Note that  $\mathbf{W}^T = \mathbf{W}$  and the second derivative is

$$\begin{aligned} \frac{\partial^2 Z(\mathbf{X}, \mathbf{X}')}{\partial \mathbf{X}_{ij} \partial \mathbf{X}_{kl}} &= - \left[ (\mathbf{W} \text{vec}(\mathbf{X}'))_{ij} \frac{\mathbf{X}'_{ij}}{\mathbf{X}_{ij}^2} + \frac{(\Delta \mathbf{X}')_{ij}}{\mathbf{X}'_{ij}} \right] \delta_{ik} \delta_{jl} \\ &\leq 0 \end{aligned} \quad (26)$$

where  $\delta_{uv}$  is defined as

$$\delta_{uv} = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{otherwise.} \end{cases} \quad (27)$$

Therefore,  $Z(\mathbf{X}, \mathbf{X}')$  is a concave function in  $\mathbf{X}$  and has a unique global maximum. The global optimum can be obtained by setting the first derivative to zero, i.e.,

$$\mathbf{X}_{ij} = \mathbf{X}'_{ij} \sqrt{\frac{(\mathbf{W} \text{vec}(\mathbf{X}'))_{ij}}{(\Delta \mathbf{X}')_{ij}}}. \quad (28)$$

Therefore, we obtain the update rule in Eq.(10) by setting  $\mathbf{X}^{(t+1)} = \mathbf{X}$  and  $\mathbf{X}^{(t)} = \mathbf{X}'$ .  $\square$

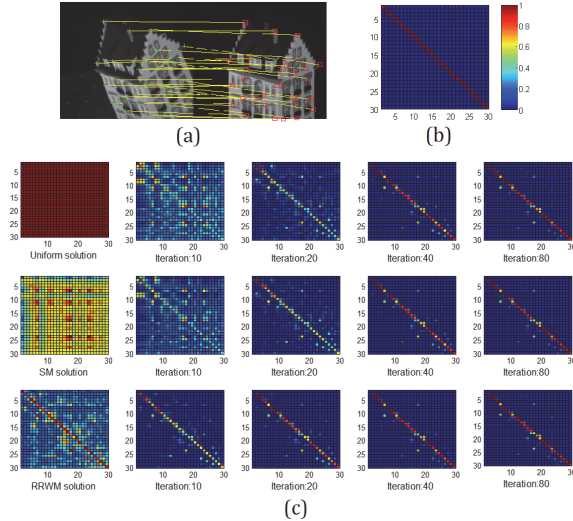


Figure 1: (a) Image corner points; (b) ground truth solution (permutation matrix); (c) solution matrix  $\mathbf{X}^{(t)}$  across different iterations with different initializations (top: start from uniform solution; middle: start from SM solution; bottom: start from RRWM solution). Note that as the iteration increases,  $\mathbf{X}^{(t)}$  approximates the ground truth permutation matrix more and more closely

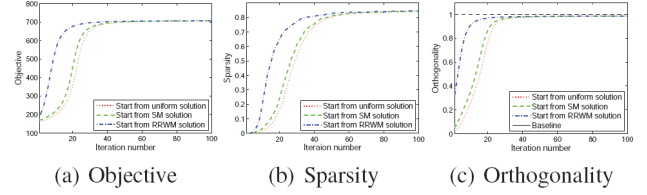


Figure 2: Objective function, sparsity and orthogonality of the solution vector  $\mathbf{X}^{(t)}$  across the iterations with different initializations under the update Eq.(10)

## Orthogonality and Approximate Discrete

As discussed before, the above multiplicative update algorithm can generate an approximate solution for NOGM problem. Here, we show experimentally that although the algorithm solves NOGM problem approximately, it can still return a desired nonnegative (approximate) orthogonal and thus sparse solution for this QAP problem. This is one important benefit of the proposed algorithm.

Figure 1(c) shows the solution  $\mathbf{X}^{(t)}$  across different iterations under the update rule Eq.(10) with three different initializations on matching two graphs generated from CMU image corner points (Figure 1(a)). The feature points and their geometric relationships correspond to the nodes and edges of the graph (see Experimental section in detail). Intuitively, we note that regardless of initialization, as the iteration increases,  $\mathbf{X}^{(t)}$  becomes more and more sparse and approximates the ground truth (permutation matrix, Figure 1(b)) more and more closely. For further illustration, we use two measurements, namely orthogonality and sparsity. For any matrix  $\mathbf{A}$ , the orthogonality is computed as follows. First, compute the normalized matrix  $\mathbf{N} = \mathbf{D}^{-1/2} \mathbf{A} \mathbf{A}^T \mathbf{D}^{-1/2}$ , where  $\mathbf{D} = \text{diag}(\mathbf{A} \mathbf{A}^T)$ . Thus, we have  $\text{diag}(\mathbf{N}) = \mathbf{I}$ . Then, we compute the orthogonality of matrix  $\mathbf{A}$  as follows,

$$\text{Orthogonality}(\mathbf{A}) = 1 - u(\mathbf{N})$$

where  $u(\mathbf{N})$  denotes the average value of the off-diagonal elements in  $\mathbf{N}$ . For any matrix  $\mathbf{A}$ , the sparsity measures the percentage of zero (close-to-zero) elements in  $\mathbf{A}$ .

Figure 2 shows the objective, sparsity and orthogonality of the solution  $\mathbf{X}^{(t)}$  across different iterations under the update algorithm Eq.(10). Here, we can observe the following: (1) Regardless of initialization, the objective of  $\mathbf{X}^{(t)}$  increases and converges after some iterations, which demonstrates the convergence of the proposed update algorithm, as discussed in Theorem 2. (2) The orthogonality of the solution  $\mathbf{X}^{(t)}$  increases and almost converges to an orthogonal (permutation) matrix at convergence. It clearly indicates the strong ability of the approximate algorithm to maintain the orthogonal constraint in the nonnegative domain, although this orthogonal constraint has been partly relaxed in the algorithm, as shown in Theorem 1. (3) The solution  $\mathbf{X}^{(t)}$  becomes more and more sparse and converges to a desired approximate discrete solution.

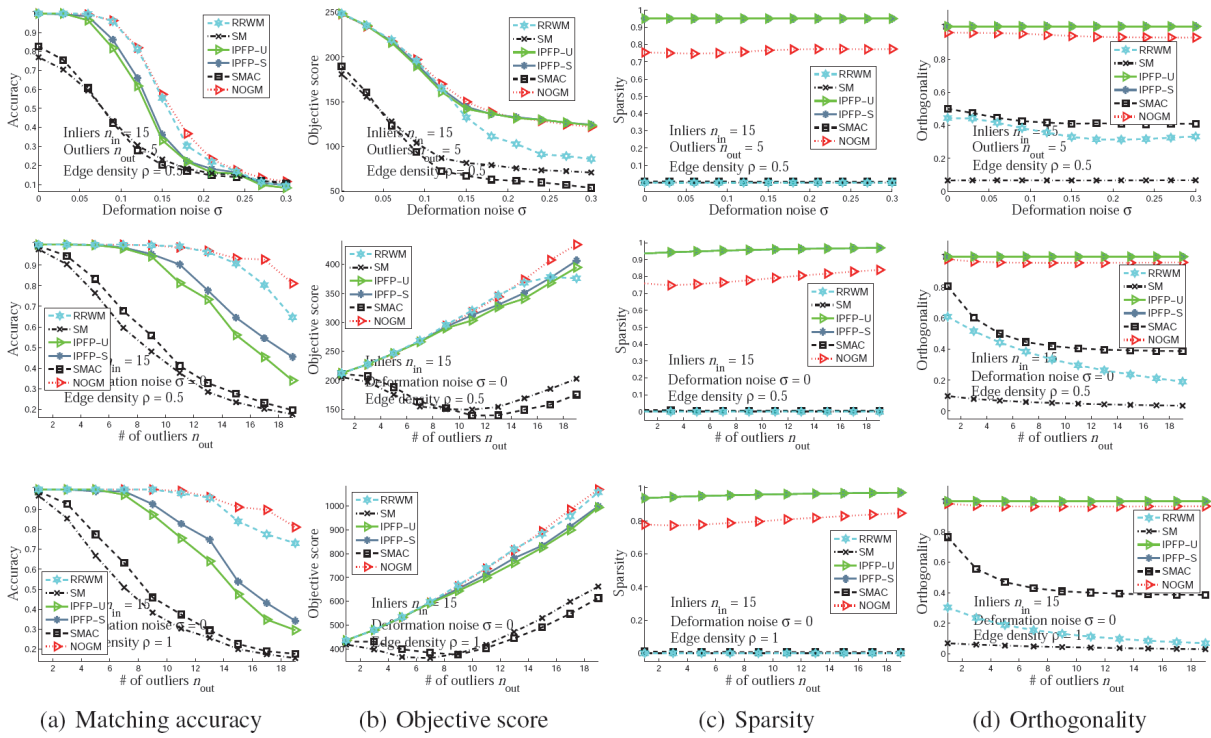


Figure 3: Comparison of graph matching results on synthetic point matching with deformation noise

## Experiments

In order to evaluate the practicality of the proposed NOGM matching method, we have applied it to several matching tasks including synthetic graph matching, feature point matching across image sequences and feature matching on real images. Our method has been compared with some recent state-of-the-art relaxation models and algorithms including SM (Leordeanu and Hebert 2005), IPFP (Leordeanu, Hebert, and Sukthankar 2009), SMAC (Cour, Srinivasan, and Shi 2006) and RRWM (Cho, Lee, and Lee 2010). We implemented IPFP with two versions: (1) IPFP-U, that is initialized by the uniform solution; (2) IPFP-S that is initialized by SM. These compared methods generally have the similar computational complexity and are most related with our NOGM algorithm.

### Synthetic data

In this section, we evaluated our NOGM algorithm on random graph matching problems. Following the experimental setting (Cho, Lee, and Lee 2010), we first generated two graphs  $G^M$  and  $G^D$ . Both of them contain  $n_{in}$  nodes. To evaluate the robustness of the our method w.r.t outlier nodes. Here, we added  $n_{out}$  outlier nodes in both  $G^M$  and  $G^D$ . For each pair of nodes, the edge is randomly generated according to the edge density  $\rho \in [0, 1]$ , and the attribute of the edge in  $G^D$  is assigned with a random value  $\mathbf{r}_{ij}^M$  distributed uniformly as  $\mathbf{r}_{ij}^M \sim U[0, 1]$ . The corresponding edge  $\mathbf{r}_{ij}^D$  in  $G^D$  was perturbed by adding a random Gaussian noise  $\sigma$  to  $\mathbf{r}_{ij}^M$ . Here,  $\sigma$  controls the level of

deformation noise. The affinity matrix  $\mathbf{W}$  is computed by  $\mathbf{W}_{ij,kl} = \exp(-\frac{\|\mathbf{r}_{ik}^D - \mathbf{r}_{jl}^M\|_F^2}{\sigma_r})$ , where scaling factor  $\sigma_r$  has been set to 0.025 in this experiment. For each noise level  $\sigma$  or  $n_{out}$ , we have generated 100 random graph pairs and then computed the average performances including matching accuracy, objective score, sparsity and orthogonality. Figure 3 summarizes the comparison results. It is noted that: (1) NOGM algorithm can always generate an approximate orthogonal and sparse solution for QAP matching problem, which demonstrates the ability of NOGM to maintain the discrete mapping constraint in its optimization. (2) NOGM generally performs better than other methods, which demonstrates the effectiveness and benefits of NOGM.

### Image sequence matching

We perform feature matching on CMU and YORK image sequences (Cho, Lee, and Lee 2010; Caetano et al. 2009; Luo, Wilson, and Hancock 2003). For each CMU hotel image, 30 landmark points have been manually marked with known correspondences. We have matched all images spaced by different frames and computed the average performances. For each image in YORK sequence, 40 landmark points have been manually marked with known correspondences. We have matched all images spaced by different frames and computed the average performances per separation gap. For each pair of images in these two sequences, we have first generated two graphs  $G^M$ ,  $G^D$  with the nodes representing the landmark points and the edges denoting the relationship (Euclidean distance) between nodes. Figure 4 summa-



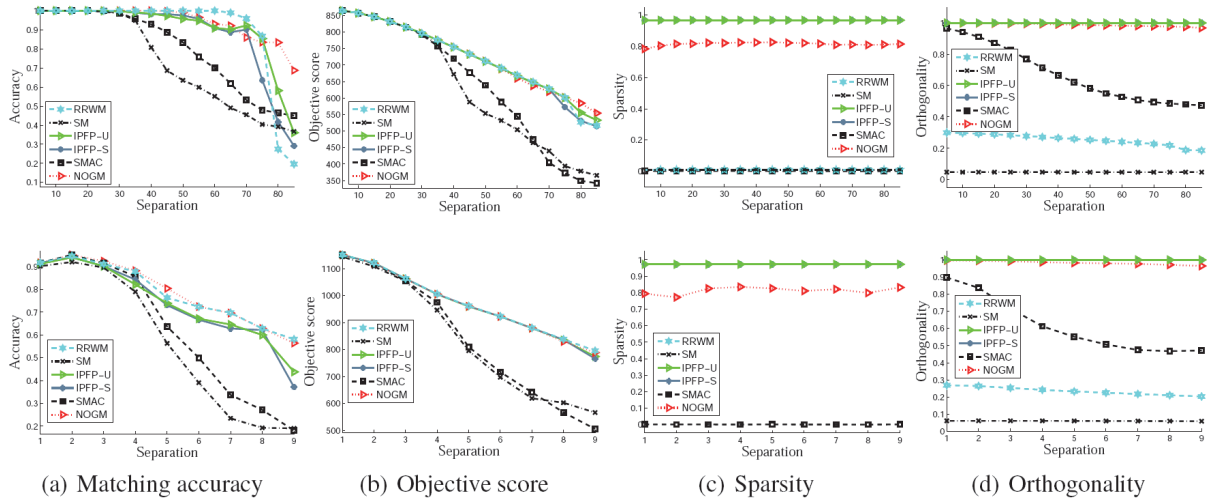


Figure 4: Comparison results across on CMU and YORK image sequences. Top: CMU sequence; Bottom: YORK sequence

izes results. We can observe that: (1) NOGM always generates approximate orthogonal and sparse solutions, which further demonstrates the sparse property of NOGM solution. (2) In CMU image dataset, as the separation exceeds 75, RRWM as well as IPFP, SM and SMAC generally break down, while our NOGM still keeps the desirable matching results. It suggests the robustness of NOGM algorithm. (3) In YORK image dataset, NOGM consistently outperforms other compared relaxation methods, indicating the effectiveness of NOGM.

## Real-world image matching

In this section, we tested our method on some real-world image feature matching tasks. Our first evaluation is conducted on the image pairs selected from Caltech-101 and MSRC datasets (Cho, Lee, and Lee 2010). Following the experimental setting (Cho, Lee, and Lee 2010), the candidate correspondences have been generated using the MSER detector and SIFT feature descriptors. Our second evaluation is performed on the image pairs (20 pairs) selected from Zurich Building Image Database (ZuBud) (Ng and Kingsbury 2010). Feature points and candidate correspondences have been detected and generated using the SIFT descriptor (Ng and Kingsbury 2010). Using the distance of SIFT descriptor, each feature in first image can match to the 8 closest features in the second image. The ground truths (correct matches) have been manually labeled for each image pair. For all image pairs in these two datasets, the average performances including true positive and false positive are computed. The comparison results are summarized in Figure 5, and some examples are shown in Figure 5. From Figure 5, we can note that NOGM algorithm returns higher true positive and lower false positive value than other comparison methods. It shows the effectiveness of the proposed NOGM method on real-world image matching tasks.

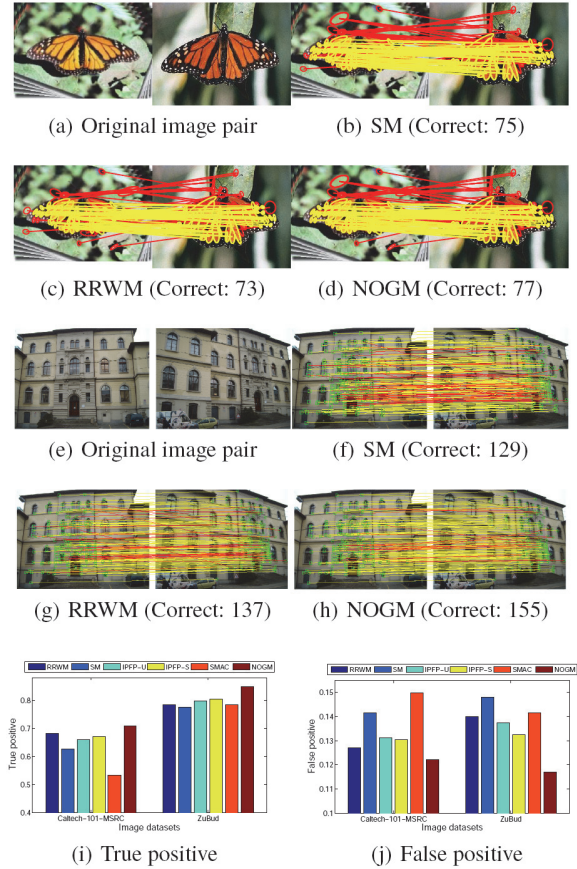


Figure 5: Matching results on the Caltech-101-MSRC and ZuBud datasets. Correct matches are marked by yellow lines, false matches by red lines

## Conclusion

NOGM aims to solve QAP problem in nonnegative orthogonal domain. NOGM can generate an optimal sparse solution for QAP problem. Promising experimental results show the effectiveness of the proposed method, which further suggests that searching for an approximate optimal solution under the nonnegative orthogonal constraint may be more important or essential for optimizing QAP problem. In our future, we will explore some more effective algorithms to solve the proposed NOGM problem.

## Acknowledgments

This work was supported by the National Key Basic Research Program of China (973 Program) (2015CB351705); National Nature Science Foundation of China (61602001, 61671018, 61472002); Natural Science Foundation of Anhui Higher Education Institutions of China (KJ2016A020); Co-Innovation Center for Information Supply & Assurance Technology, Anhui University; the Open Projects Program of National Laboratory of Pattern Recognition.

## References

- Adamczewski, K.; Suh, Y.; and Lee, K. M. 2015. Discrete tabu search for graph matching. In *IEEE International Conference on Computer Vision*, 109–117.
- Caetano, T. S.; McAuley, J. J.; Cheng, L.; Le, Q. V.; and Smola, A. J. 2009. Learning graph matching. *PAMI* 31(6):1048–1058.
- Chen, G.; Wang, F.; and Zhang, C. 2007. Collaborative filtering using orthogonal nonnegative matrix tri-factorization. In *ICDM Workshops*, 303–308.
- Cho, M.; Lee, J.; and Lee, K. M. 2010. Reweighted random walks for graph matching. In *ECCV*, 492–505.
- Choi, O., and Kweon, I. S. 2009. Robust feature point matching by preserving local geometric consistency. *CVIU* 113(6):726–742.
- Conte, D.; Foggia, P.; Sansone, C.; and Vento, M. 2004. Thirty years of graph matching in pattern recognition. *IJPRAI* 18(3):265–298.
- Cour, M.; Srinivasan, P.; and Shi, J. 2006. Balanced graph matching. In *NIPS*, 313–320.
- Ding, C.; Li, T.; Peng, W.; and Park, H. 2006. Orthogonal nonnegative matrix t-factorizations for clustering. In *KDD*, 126–135.
- Ding, C.; Li, T.; and Jordan, M. I. 2008. Nonnegative matrix factorization for combinatorial optimization: Spectral clustering, graph matching and clique finding. In *ICDM*, 183–192.
- Ding, C.; Li, T.; and Jordan, M. I. 2010. Convex and semi-nonnegative matrix factorization. *PAMI* 32(1):45–55.
- Enqvist, O.; Josephson, K.; and Kahl, F. 2009. Optimal correspondences from pairwise constraints. In *ICCV*, 1295–1302.
- Gold, S., and A, R. 1996. A graduated assignment algorithm for graph matching. *PAMI* 18(4):377–388.
- Jiang, B.; Tang, J.; Ding, C.; and Luo, B. 2015. A local sparse model for matching problem. In *AAAI*, 3790–3796.
- Jiang, B.; Tang, J.; Cao, X.; and Luo, B. 2017. Lagrangian relaxation graph matching. *Pattern Recognition* 61:255–265.
- Lee, D. D., and Seung, H. S. 2001. Algorithms for nonnegative matrix factorization. In *NIPS*, 556–562.
- Leordeanu, M., and Hebert, M. 2005. A spectral technique for correspondence problem using pairwise constraints. In *ICCV*, 1482–1489.
- Leordeanu, M.; Hebert, M.; and Sukthankar, R. 2009. An integer projected fixed point method for graph matching and map inference. In *NIPS*, 1114–1122.
- Li, H.; Adali, T.; Wang, W.; Emge, D.; and Cichocki, A. 2007. Non-negative matrix factorization with orthogonality constraints and its applications to reman spectroscopy. *Journal of VLSI Signal Processing* 48:83–97.
- Liu, Z. Y.; Qiao, H.; and Xu, L. 2012. An extended path following algorithm for graph matching problem. *PAMI* 34(7):1451–1456.
- Luo, B.; Wilson, R. C.; and Hancock, E. R. 2003. Spectral embedding of graphs. *Pattern Recognition* 36(10):2213–2230.
- Ng, E. S., and Kingsbury, N. G. 2010. Matching of interest point groups with pairwise spatial constraints. In *ICIP*, 2693–2696.
- van Wyk, B. J., and van Wyk, M. A. 2004. A pocs-based graph matching algorithm. *PAMI* 16(11):1526–1530.
- Zaslavskiy, M.; Bach, F.; and Vert, J. P. 2009. A path following algorithm for the graph matching problem. *PAMI* 31(12):2227–2242.
- Zhang, Z.; Shi, Q.; McAuley, J.; Wei, W.; Zhang, Y.; and Hengel, A. V. D. 2016. Pairwise matching through max-weight bipartite belief propagation. In *IEEE Conference on Computer Vision and Pattern Recognition*.
- Zhou, F., and la Torre, F. D. 2012. Factorized graph matching. In *CVPR*, 127–134.