

---

# Near-Optimal Joint Object Matching via Convex Relaxation

---

**Yuxin Chen**

Department of Electrical Engineering, Stanford University, Stanford, CA 94305, USA

YXCHEN@STANFORD.EDU

**Leonidas Guibas**

Department of Computer Science, Stanford University, Stanford, CA 94305, USA

GUILAS@CS.STANFORD.EDU

**Qixing Huang**

Department of Computer Science, Stanford University, Stanford, CA 94305, USA

HUANGQX@STANFORD.EDU

## Abstract

Joint object matching aims at aggregating information from a large collection of similar instances (e.g. images, graphs, shapes) to improve the correspondences computed between pairs of objects, typically by exploiting global map compatibility. Despite some practical advances on this problem, from the theoretical point of view, the error-correction ability of existing algorithms are limited by a constant barrier — none of them can provably recover the correct solution when more than a constant fraction of input correspondences are corrupted. Moreover, prior approaches focus mostly on fully similar objects, while it is practically more demanding and realistic to match instances that are only partially similar to each other.

In this paper, we propose an algorithm to jointly match multiple objects that exhibit only partial similarities, where the provided pairwise feature correspondences can be densely corrupted. By encoding a consistent partial map collection into a 0-1 semidefinite matrix, we attempt recovery via a two-step procedure, that is, a spectral technique followed by a parameter-free convex program called MatchLift. Under a natural randomized model, MatchLift exhibits near-optimal error-correction ability, i.e. it guarantees the recovery of the ground-truth maps even when a dominant fraction of the inputs are randomly corrupted. We evaluate the proposed algorithm on various benchmark data sets including synthetic examples and real-world examples, all of which confirm the practical applicability of the proposed algorithm.

## 1. Introduction

Finding consistent relations across multiple objects is a fundamental scientific problem spanning many fields. A partial list includes jigsaw puzzle solving (Cho et al., 2010), structure from motion (Zach et al., 2010), and reassembly of fragmented objects and documents (Huang et al., 2006), to name just a few. Compared with the rich literature in pairwise matching (e.g., of graphs, images or shapes), joint matching of multiple objects has not been well explored. A naive approach for joint object matching is to pick a base object and perform *pairwise matching* with each of the remaining objects, effectively using the base object as a reference when comparing any two objects. However, as pairwise methods typically generate noisy results due to unavoidable matching ambiguities and other errors, the performance of such approaches is often far from satisfactory in practice. This gives rise to the question as to how to aggregate and exploit information from multiple maps that one might compute across several object pairs, in order to improve the consistency and quality of global object matching, ideally to recover the ground-truth maps.

In this paper, we represent each object as a set of points or elements, and investigate the problem of joint matching over  $n$  different sets reflecting the same universe, for which the input / observation is a collection of pairwise maps computed in isolation. A natural and popular criterion for examining the global matching compatibility is called cycle-consistency or path invariance, namely, composition of maps between two objects should be independent of the connecting path chosen. Such a criterion has recently been invoked in many algorithms (Roberts et al., 2011; Zach et al., 2010; Nguyen et al., 2011; Huang et al., 2012; Kim et al., 2012) to detect outliers among the noisy input maps. These works have demonstrated experimentally that one can use inconsistent cycles to prune corrupted maps, provided that the corruption rate is sufficiently small.

Despite these empirical advances, little is known on the theoretical side — the conditions under which the ground-

truth maps can be reliably recovered from the noisy inputs. Recent work by (Huang & Guibas, 2013) presented the first theoretical guarantee for robust joint matching, and (Pachauri et al., 2013) studied the performance of spectral methods for joint matching. However, there are several fundamental and practical challenges that remain unaddressed.

- **Dense Input Errors.** The state-of-the-art results do not provide theoretical guarantees when more than 50% of the inputs are corrupted. This gives rise to a natural question regarding their applicability in the presence of highly noisy sources, in which case the majority of the input maps might be corrupted. Observe that, as the number  $n$  of objects to be matched increases, the number of pairwise maps one can obtain grows quadratically with  $n$ . As a result, dense error correction becomes information theoretically possible when a large collection of objects are present, as long as the global map consistency can be appropriately exploited. However, the challenge remains as to whether there exist *computationally feasible* methods that can provably detect and remove dense outliers.

- **Partial Similarity.** To the best of our knowledge, all prior approaches deal only with a restricted scenario, where the ground-truth maps are given by full isomorphisms (i.e. *one-to-one* correspondences between any two sets). In reality, a collection of objects typically exhibit only partial similarity, as in the case of images of the same scene but from different camera positions, where different occlusions are usually present in the images. These practical scenarios require consistent matching of multiple objects that are only partially similar to each other.

- **Incomplete Input Maps.** Computing pairwise maps between all object pairs is often expensive, sometimes infeasible, and in fact unnecessary. Depending on the characteristics of the input sources, one might be able to infer unobserved maps from a small set of noisy pairwise matches. While (Huang & Guibas, 2013) considered incomplete inputs, the tradeoff between the undersampling factor and the error-correction ability remains unexplored.

All in all, practical applications require matching partially similar objects from a small fraction of densely corrupted pairwise maps — a goal this paper aims to achieve.

## 1.1. Contributions

- 1) **Algorithms:** Inspired by the recent evidence on the power of convex relaxation, we propose to solve the joint matching problem via a semidefinite program called MatchLift. The algorithm relaxes the binary-value constraints, and attempts to maximize the compatibility between the input and the recovered maps. The program is based on a semidefinite conic constraint that depends on the total number  $m$  of distinct elements to be matched. To this end, we propose to pre-estimate  $m$  via a spectral method. Our methodology is essentially *parameter free*, and can be solved by scalable optimization algorithms.

- 2) **Theory:** We derive performance guarantees for exact matching. Somewhat surprisingly, MatchLift admits perfect map recovery even in the presence of dense input corruption. Our findings reveal the near-optimal error-correction ability of MatchLift, i.e. as  $n$  grows, the algorithm is guaranteed to work even when almost all inputs behave as random outliers. Besides, while the presence of partial similarity inevitably incurs more severe types of input errors, MatchLift exhibits a strong recovery ability nearly equivalent to that in the full-similarity scenario. Finally, in many situations, MatchLift succeeds even with minimal input complexity, in the sense that it can reliably fill in all unobserved maps based on very few noisy partial inputs, as soon as the provided maps form a connected graph. This is information theoretically optimal.

- 3) **Applicability:** We have evaluated the performance of MatchLift on several benchmark datasets. These datasets include several synthetic examples as well as real examples from several popular benchmarks. Experimental results on synthetic examples justify our theoretical performance guarantees. On real datasets, the quality of the maps generated by MatchLift outperforms the state-of-the-art object matching and graph clustering algorithms.

## 1.2. Related Work

**Object Matching.** Early work on object matching focused primarily on matching pairs of objects in isolation (e.g. (Cour et al., 2007)). Due to the limited information and ambiguities present in an isolated object pair, pairwise matching techniques can easily, sometimes unavoidably, generate false correspondences. The last few years have witnessed a flurry of activity in joint object matching, e.g. (Kim et al., 2012; Huang et al., 2012; Huang & Guibas, 2013), which exploited the cycle-consistency criterion to prune outliers. Our fundamental understanding has recently been advanced by (Huang & Guibas, 2013). However, none of the prior works have provided recovery guarantees when the majority of input correspondences are outliers, nor were they able to accommodate practical scenarios where objects only exhibit partial similarity. In a recent work (Pachauri et al., 2013), the authors developed a spectral-based approach in the full similarity case. Although this method provides theoretical analyses, the errors considered therein are modeled as Gaussian-Wigner noise, which is not realistic in our setting.

**Graph Clustering and Synchronization.** The joint matching problem can be treated as a structured graph clustering (GC) problem, where graph nodes represent points on objects and the edge set encodes all correspondences. In this regard, each GC algorithm (Bansal et al., 2004; Jalali et al., 2011; Chen et al., 2012; Jalali & Srebro, 2012) provides a heuristic for graph matching. Nevertheless, there are several intrinsic structural properties in our setting that are not explored by generic GC approaches. First, our input takes a block-matrix form, where each block is highly

*structured* (i.e. doubly-substochastic), *sparse*, and inter-dependent. Second, the points belonging to the same object are *mutually exclusive* with each other. As a result, the findings for generic GC methods do not deliver encouraging guarantees when applied to our setting. Finally, we note that our joint matching problem is very relevant to various synchronization problems (Wang & Singer, 2013; Chaudhury et al., 2013; Bandeira et al., 2013; Abbe et al., 2014) and the insightful approaches adopted therein.

## 2. Problem Formulation and Preliminaries

### 2.1. Terminology

**Set.** We represent the objects to be matched as discrete sets. For example, these sets encode feature points when matching images or shapes.

**Partial Map.** Given two discrete sets  $\mathcal{S}$  and  $\mathcal{S}'$ , a subset  $\phi \subset \mathcal{S} \times \mathcal{S}'$  is termed a partial map if each element of  $\mathcal{S}$  (resp.  $\mathcal{S}'$ ) is paired with *at most one* element of  $\mathcal{S}'$  (resp.  $\mathcal{S}$ ) — in particular, not all elements need to be paired.

**Map Graph.** A graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  is called a map graph w.r.t.  $n$  sets  $\mathcal{S}_1, \dots, \mathcal{S}_n$  if (i)  $\mathcal{V} := \{\mathcal{S}_1, \dots, \mathcal{S}_n\}$ , and (ii)  $(\mathcal{S}_i, \mathcal{S}_j) \in \mathcal{E}$  implies that pairwise estimates on the partial maps  $\phi_{ij}$  and  $\phi_{ji}$  between  $\mathcal{S}_i$  and  $\mathcal{S}_j$  are available.

### 2.2. Input and Output

The input and expected output for the joint object matching problem are described as follows.

**Input (Noisy Pairwise Maps).** Given  $n$  sets  $\mathcal{S}_1, \dots, \mathcal{S}_n$  with respective cardinality  $m_1, \dots, m_n$  and a (possibly sparse) map graph  $\mathcal{G}$ , the input to the recovery algorithm consists of partial maps  $\phi_{ij}^{\text{in}}$  ( $(i, j) \in \mathcal{G}$ ) between  $\mathcal{S}_i$  and  $\mathcal{S}_j$  estimated in isolation, using any off-the-shelf pairwise matching method. Note that the input maps  $\phi_{ij}^{\text{in}}$  one obtain might not agree, partially or totally, with the ground truth.

**Output (Consistent Global Matching).** The main objective of this paper is to detect and prune incorrect pairwise input maps in an efficient and reliable manner. Specifically, we aim at proposing a tractable algorithm that returns a full collection of partial maps  $\{\phi_{ij} \mid 1 \leq i, j \leq n\}$  that are provably the ground-truth maps under mild conditions.

As will be detailed later, the key idea of our approach is to explore global consistency across all pairwise maps. We will introduce a novel convex relaxation tailored to the structure of the input maps (Section 3) and investigate its theoretical performance (Section 4).

### 2.3. Joint Matching in Matrix Form

In the same spirit as most convex relaxation and spectral techniques (e.g., (Huang & Guibas, 2013; Pachauri et al., 2013)), we use 0-1 matrices to encode point-to-point maps across objects. Specifically, we encode a partial map

$\phi_{ij} : \mathcal{S}_i \mapsto \mathcal{S}_j$  as a binary matrix  $\mathbf{X}_{ij} \in \{0, 1\}^{|\mathcal{S}_i| \times |\mathcal{S}_j|}$  such that  $\mathbf{X}_{ij}(s, s') = 1$  iff  $(s, s') \in \phi_{ij}$ . As each map associates each vertex with at most one corresponding vertex, the map matrices  $\mathbf{X}_{ij}$  shall satisfy the following doubly sub-stochastic constraints:

$$\mathbf{0} \leq \mathbf{X}_{ij}\mathbf{1} \leq \mathbf{1}, \quad \mathbf{0} \leq \mathbf{X}_{ij}^\top\mathbf{1} \leq \mathbf{1}. \quad (1)$$

We then use an  $n \times n$  block matrix  $\mathbf{X} \in \{0, 1\}^{N \times N}$  to encode the entire collection of partial maps  $\{\phi_{ij} \mid 1 \leq i, j \leq n\}$  over  $\{\mathcal{S}_1, \dots, \mathcal{S}_n\}$ :

$$\mathbf{X} = \begin{pmatrix} \mathbf{I}_{m_1} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1n} \\ \mathbf{X}_{21} & \mathbf{I}_{m_2} & \cdots & \mathbf{X}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{n1} & \cdots & \cdots & \mathbf{I}_{m_n} \end{pmatrix}, \quad (2)$$

where  $m_i := |\mathcal{S}_i|$  and  $N := \sum_{i=1}^n m_i$ . Note that diagonal blocks are identity matrices, representing self maps.

For notational simplicity, we will use the binary map matrices  $\mathbf{X}_{ij}^{\text{in}} \in \{0, 1\}^{m_i \times m_j}$  throughout to denote the collection of pairwise input maps

## 3. Methodology

### 3.1. MatchLift: A Novel Two-Step Algorithm

We start by discussing the consistency constraint on the underlying ground-truth maps. Assume that there exists a universe  $\mathcal{S} = \{1, \dots, m\}$  of  $m$  elements such that i) each object  $\mathcal{S}_i$  is a (partial) image of  $\mathcal{S}$ ; ii) each element in  $\mathcal{S}$  is contained in at least one object  $\mathcal{S}_i$ . Then the ground-truth correspondences shall connect those points across objects that represent the same element.

Formally speaking, let the binary matrices  $\mathbf{Y}_i \in \{0, 1\}^{m_i \times m}$  encode the underlying correspondence between object  $i$  and the universe, i.e.  $\mathbf{Y}_i(s_i, s) = 1$  iff  $s_i \in \mathcal{S}_i$  represents  $s \in \mathcal{S}$ . This way we can express  $\mathbf{X} = \mathbf{Y}\mathbf{Y}^\top$  with  $\mathbf{Y} = (\mathbf{Y}_1^\top, \dots, \mathbf{Y}_n^\top)^\top$ , which makes clear that  $\text{rank}(\mathbf{X}) = m$ . This corresponds to the graph partitioning setting with  $m$  cliques. Consequently, a natural candidate is to seek a low-rank and positive semidefinite (PSD) matrix to approximate the input. However, this property itself is not effective enough in exploiting the structure underlying the map collection.

A more powerful formulation arises from the observation that even under dense input corruption, we are often able to obtain reliable estimates on  $m$  — the universe size, using spectral techniques. This motivates us to incorporate the information of  $m$  into the formulation so as to develop tighter relaxation. Specifically, we lift  $\mathbf{X}$  with one more dimension and consider

$$\begin{bmatrix} m & \mathbf{1}^\top \\ \mathbf{1} & \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{1}^\top \\ \mathbf{Y} \end{bmatrix} \begin{bmatrix} \mathbf{1} & \mathbf{Y}^\top \end{bmatrix} \succeq \mathbf{0}. \quad (3)$$

By Schur complement condition, this is equivalent to  $\mathbf{X} - \frac{1}{m}\mathbf{1}\mathbf{1}^\top \succeq \mathbf{0}$ , which is strictly tighter than  $\mathbf{X} \succeq \mathbf{0}$ . Intu-

itively, the formulation (3) entitles us one extra degree of freedom to assist in outlier pruning. More precisely, this degree of freedom is approximately along the direction of  $\mathbf{1}\mathbf{1}^\top$ , which is crucial in globally “debiasing” the errors. We now present our two-step procedure as follows.

**Step I: Estimating  $m$ .** We estimate  $m$  by tracking the spectrum of the input  $\mathbf{X}^{\text{in}}$ . According to common wisdom (e.g. (Keshavan et al., 2010)), the input matrix  $\mathbf{X}^{\text{in}}$  needs to be pre-trimmed in order to remove the undesired bias effect caused by over-represented rows / columns. One candidate trimming procedure is provided as follows.

- **Trimming Procedure.** Set  $d_{\min}$  to be the smallest vertex degree of  $\mathcal{G}$ , and we say the  $i$ th ( $1 \leq i \leq n$ ) vertex is *over-represented* if the degree of  $i$  in  $\mathcal{G}$  exceeds  $2d_{\min}$ . Then for each overrepresented vertex  $i$ , randomly sample  $2d_{\min}$  edges incident to it and set to zero all blocks  $\mathbf{X}_{ij}^{\text{in}}$  associated with the remaining edges.

With this pre-trimming procedure, our algorithm for estimating  $m$  proceeds as described in Algorithm 1.

#### Algorithm 1 Estimating the size $m$ of the universe $\mathcal{S}$

- 1) trim  $\mathbf{X}^{\text{in}}$ , and let  $\tilde{\mathbf{X}}^{\text{in}}$  be the output.
- 2) perform eigen-decomposition of  $\tilde{\mathbf{X}}^{\text{in}}$ ; denote by  $\lambda_i$  the  $i$ th largest eigenvalue.
- 3) **output:**  $\hat{m} := \arg \max_{M \leq i < N} |\lambda_i - \lambda_{i+1}|$ , where  $M = \max\{2, \max_{1 \leq i \leq n} m_i\}$ .

In short, Algorithm 1 returns an estimate of  $m$  via spectral methods, which outputs the number of dominant principal components of  $\mathbf{X}^{\text{in}}$ .

**Step II: Map Recovery.** Now that we have obtained an estimate on  $m$ , we are in position to present our optimization heuristic that exploits the property (3). In order to guarantee that the recovery is close to the provided maps  $\phi_{ij}^{\text{in}}$ , one alternative is to maximize correspondence agreement (i.e. the number of compatible non-zero entries) between the input and output. This results in an objective function:  $\sum_{(i,j) \in \mathcal{G}} \langle \mathbf{X}_{ij}^{\text{in}}, \mathbf{X}_{ij} \rangle$ . Additionally, since a *non-negative* map matrix  $\mathbf{X}$  is inherently sparse, it is natural to add an elementwise  $\ell_1$  regularization term to encourage sparsity, which in our case reduces to  $\|\mathbf{X}\|_1 = \langle \mathbf{1} \cdot \mathbf{1}^\top, \mathbf{X} \rangle$ .

Since searching over all 0-1 map matrices is intractable, we propose to relax the binary constraints. This leads to the following semidefinite program referred to as *MatchLift*:

$$\begin{aligned}
 (\text{MatchLift}) \quad & \underset{\mathbf{X} \in \mathbb{R}^{N \times N}}{\text{maximize}} \quad \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{X}_{ij}^{\text{in}}, \mathbf{X}_{ij} \rangle - \lambda \langle \mathbf{1} \cdot \mathbf{1}^\top, \mathbf{X} \rangle \\
 \text{subject to} \quad & \mathbf{X}_{ii} = \mathbf{I}_{m_i}, \quad 1 \leq i \leq n, \\
 & \mathbf{X} \geq \mathbf{0}, \\
 & \begin{bmatrix} m & \mathbf{1}^\top \\ \mathbf{1} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}.
 \end{aligned}$$

**Remark.** Here,  $\lambda$  represents the regularization parameter that balances the agreement to input correspondences and the sparsity structure. As we will show, MatchLift is not sensitive to the choice of  $\lambda$ . By default, one can set  $\lambda = \sqrt{|\mathcal{E}|}/(2n)$ , which results in a *parameter-free* formulation.

This algorithm, all at once, attempts to disentangle the ground truth and outliers as well as predict unobserved maps via convex relaxation, inspired by recent success in sparse and low-rank matrix decomposition (Candès et al., 2011). Since the ground truth matrix is *simultaneously* low-rank and sparse; existing methodologies, which focus on dense low-rank matrices, typically yield loose, uninformative bounds in our setting.

#### 3.2. Alternating Direction Methods of Multipliers (ADMM)

Most advanced off-the-shelf SDP solvers like SeDuMi or MOSEK are typically based on interior point methods, and such second-order methods are unable to handle problems with large dimensionality. For practical applicability, we propose a scalable first-order optimization algorithm for approximately solving MatchLift, which is a variant of the ADMM method for semidefinite programs presented in (Wen et al., 2010). Theoretically it is guaranteed to converge. Empirically, it is often the case that ADMM converges to modest accuracy within a reasonable amount of time, and produces desired results with the assistance of appropriate rounding procedures. This feature makes ADMM practically appealing in our case since the ground-truth matrix is known to be a 0-1 matrix, for which moderate entry-wise precision is sufficient to ensure good rounding accuracy. The algorithm is detailed in (Chen et al., 2014) and the supplemental materials.

#### 3.3. Rounding Strategy

As Matchlift solves a relaxed program of the original convex problem, it may return fractional solutions. In this case, we propose a greedy rounding method to generate valid partial maps. Given the solution  $\hat{\mathbf{X}}$  to MatchLift, the proposed strategy proceeds as in Algorithm 2. One can verify that this simple deterministic rounding strategy returns a matrix that encodes a consistent collection of partial maps. Note that  $\mathbf{v}_i^\top$  denotes the  $i$ th row of a matrix  $\mathbf{V}$ .

#### 4. Theoretical Guarantees: Exact Recovery

Our heuristic algorithm MatchLift recovers, under a natural randomized setting, the ground-truth maps even when only a vanishing portion of the input correspondences are correct. Furthermore, MatchLift succeeds with minimal input complexity, namely, the algorithm is guaranteed to work as soon as those input maps that coincide with the ground truth maps form a connected map graph.

**Algorithm 2** Rounding Strategy

**initialize** compute the top  $r$  eigenvalues  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$  and eigenvectors  $\mathbf{U} = (\mathbf{u}_1, \dots, \mathbf{u}_r)$  of  $\hat{\mathbf{X}}$ , where  $r$  is an estimate of the total number distinctive points to be recovered. Form  $\mathbf{V} = \mathbf{U}\Sigma^{\frac{1}{2}}$ .

**repeat**

1) Let  $\mathbf{O}$  be a unitary matrix that obeys  $\mathbf{O}\mathbf{v}_1 = \mathbf{e}_1$ , and set  $\mathbf{V} \leftarrow \mathbf{V}\mathbf{O}^\top$ .

2) For each of the remaining rows  $\mathbf{v}_i$  belonging to each set  $\mathcal{S}_j$  ( $i \in \mathcal{S}_j$ ), perform

$$\mathbf{v}_i \leftarrow \mathbf{e}_1, \text{ if } \langle \mathbf{v}_i, \mathbf{v}_1 \rangle > 0.5. \text{ and } i = \arg \max_{l \in \mathcal{S}_j} \langle \mathbf{v}_l, \mathbf{v}_1 \rangle$$

3) All indices  $i$  obeying  $\mathbf{v}_i = \mathbf{e}_1$  are declared to be matched with each other, and are then removed. Repeat 1) for the next row that has not been fixed.

**until** all the rows of  $\mathbf{V}$  have been fixed.

#### 4.1. Randomized Model

In the following, we present a natural noisy model, under which the capability of MatchLift is easiest to interpret.

**Randomized Model.** Consider a universe  $[m] := \{1, 2, \dots, m\}$ . The randomized setting consider herein is generated through the following procedure.

- For each set  $\mathcal{S}_i$  ( $1 \leq i \leq n$ ), each point  $s \in [m]$  is included in  $\mathcal{S}_i$  independently with probability  $p_{\text{set}}$ .
- Each  $\mathbf{X}_{ij}^{\text{in}}$  is observed / computed independently with probability  $p_{\text{obs}}$ .
- Each observed  $\mathbf{X}_{ij}^{\text{in}}$  coincides with the ground truth independently with probability  $p_{\text{true}} = 1 - p_{\text{false}}$ .
- Each *observed but incorrect*  $\mathbf{X}_{ij}^{\text{in}}$  is independently drawn from a set of partial map matrices satisfying<sup>1</sup>

$$\mathbb{E}\mathbf{X}_{ij}^{\text{in}} = \mathbf{1} \cdot \mathbf{1}^\top / m, \text{ if } \mathbf{X}_{ij}^{\text{in}} \text{ is observed and corrupted.}$$

#### 4.2. Main Theorem: Near-Optimal Matching

We are now in position to state our main results, which provide theoretical performance guarantees for our algorithms. All proofs of the main results are deferred to (Chen et al., 2014) and the supplemental materials.

**Theorem 1 (Accurate Estimation of  $m$ ).** Consider the above randomized model. There exists an absolute constant  $c_1 > 0$  such that with probability exceeding  $1 - \frac{1}{m^5 n^5}$ , the estimate on  $m$  returned by Algorithm 1 is exact as long as

$$p_{\text{true}} \geq c_1 \log^2(mn) / \sqrt{n p_{\text{obs}} p_{\text{set}}}. \quad (4)$$

<sup>1</sup>This holds, for example, when the augmented block (i.e. that obtained by enhancing  $\mathcal{S}_i$  and  $\mathcal{S}_j$  to have all  $m$  elements) is drawn from the set of permutation matrices uniformly at random. Note, however, that this assumption can be significantly relaxed without degrading the matching performance.

Theorem 1 ensures that one can obtain perfect estimate on the universe size or, equivalently, the rank of the ground truth map matrix via spectral methods. With accurate information on  $m$ , MatchLift allows perfect matching from densely corrupted inputs, as revealed below.

**Theorem 2 (Exact and Robust Matching).** Consider the randomized model described above. There exist universal constants  $c_0, c_1, c_2 > 0$  such that for any

$$c_1 \left( \frac{p_{\text{obs}}}{m} + \sqrt{\frac{p_{\text{obs}} \log(mn)}{np_{\text{set}}^3}} \right) \leq \lambda \leq c_2 \sqrt{p_{\text{obs}} \log(mn)},$$

if the non-corruption rate obeys

$$p_{\text{true}} > c_0 \log^2(mn) / \sqrt{n p_{\text{obs}} p_{\text{set}}^2}, \quad (5)$$

then the solution to MatchLift is exact and unique with probability exceeding  $1 - (mn)^{-3}$ .

Note that the performance is not sensitive to  $\lambda$  as it can be arbitrarily chosen within a large range. The implications of Theorem 2 are summarized as follows.

- a) **Near-Optimal Recovery under Dense Errors.** Under the randomized model, MatchLift succeeds in pruning all outliers and recovering the ground truth with high probability. Somewhat surprisingly, this is guaranteed to work even when the non-corrupted pairwise maps account for only a *vanishing* fraction of the inputs. As a result, MatchLift achieves near-optimal recovery ability in the sense that, as the number  $n$  of objects grows, the outlier-tolerance rate it can achieve can be arbitrarily close to 1. Equivalently speaking, in the high-dimensional regime, almost all input maps can be badly corrupted by random errors without degrading the matching accuracy. This in turn highlights the significance of joint object matching: no matter how noisy the input sources are, perfect matching can be obtained as long as sufficiently many instances are available.

To the best of our knowledge, none of the prior results can support perfect recovery with more than 50% corruptions, regardless of how large  $n$  can be. The only comparative performance is reported for the robust PCA setting, where semidefinite relaxation enables dense error correction (Chen et al., 2013). However, their conditions do not apply in our case. Experimentally, applying RPCA on joint matching is unable to tolerate dense errors, as reported in Section 5.

- b) **Exact Matching of Partially Similar Objects.** The challenge for matching partially similar objects arises in that the overlapping ratio between each pair of objects is in the order of  $p_{\text{set}}^2$  while the size of each object is in the order of  $p_{\text{set}}$ . As correct correspondences only come from overlapping regions, it is expected that with a fixed  $p_{\text{false}}$ , the matching ability degrades when  $p_{\text{set}}$

decreases, which is reflected by (5). Note that the order of the fault-tolerance rate with  $n$  is independent of  $p_{\text{set}}$  as long as  $p_{\text{set}}$  is bounded away from 0.

- c) **Minimal Input Complexity.** Suppose that  $p_{\text{set}}$  and  $p_{\text{false}}$  are both constants bounded away from 0 and 1, and that  $m = n^{O(\text{poly log}(n))}$ . Condition (5) asserts that: the algorithm is able to separate outliers and fill in all missing maps reliably with no errors, as soon as the input complexity (i.e. the number of pairwise maps provided) is about the order of  $n \text{poly log}(n)$ . Recall that the connectivity threshold for an Erdős–Renyi graph  $\mathcal{G}(n, p_{\text{obs}})$  is  $p_{\text{obs}} > \log n/n$ . This implies that MatchLift allows exact recovery nearly as soon as the input complexity exceeds the connectivity threshold.

Finally, if the universe size  $m$  does not scale with  $n$ , then it has been shown (Chen & Goldsmith, 2014) that no method whatsoever can recover the ground truth if  $p_{\text{true}} < O(1/\sqrt{np_{\text{obs}}})$ . That said, MatchLift achieves the information theoretically optimal error-correction ability except for some logarithmic factor under our randomized model.

### 4.3. Comparison with Prior Approaches

Our exact recovery condition significantly outperforms the best-known performance guarantees, including various SDP heuristics for matching problems, as well as generic graph clustering approaches when applied to object matching, detailed below.

**Semidefinite Programming:** The SDP formulation proposed for synchronization problems (Wang & Singer, 2013) asymptotically admits exact recovery in the *full-similarity* setting when  $p_{\text{true}} > c_1$  for some absolute constant  $c_1 \approx 50\%$  (except for the binary case (Abbe et al., 2014)). One might also attempt recovery via robust PCA (Candès et al., 2011). In order to enable dense error correction, robust PCA requires the sparse components (which is  $\mathbf{X}^{\text{in}} - \mathbf{X}^{\text{gt}}$  here with  $\mathbf{X}^{\text{gt}}$  denoting the ground truth) to exhibit random signs (Chen et al., 2013). This cannot be satisfied here since the sign of  $\mathbf{X}^{\text{in}} - \mathbf{X}^{\text{gt}}$  is highly biased.

**Graph Clustering:** Various graph clustering approaches have been proposed with theoretical guarantees under different randomized settings (Jalali et al., 2011; Jalali & Srebro, 2012). These results typically operate under the assumption that in-cluster and inter-cluster correspondences are independently corrupted, which does not apply in our model. Due to the block structure input model, these two types of corruptions are highly correlated and usually exhibit order-of-magnitude difference in corruption rate. To facilitate the comparison, we evaluate the most recent deterministic guarantees obtained by Jalali et al (Jalali & Srebro, 2012). Simple calculation reveals that the guarantee in (Jalali & Srebro, 2012) requires  $p_{\text{true}} > \frac{m}{m+1}$ , which does not deliver encouraging guarantees compared with  $p_{\text{true}} > \Theta(1/\sqrt{n})$  achieved by MatchLift.

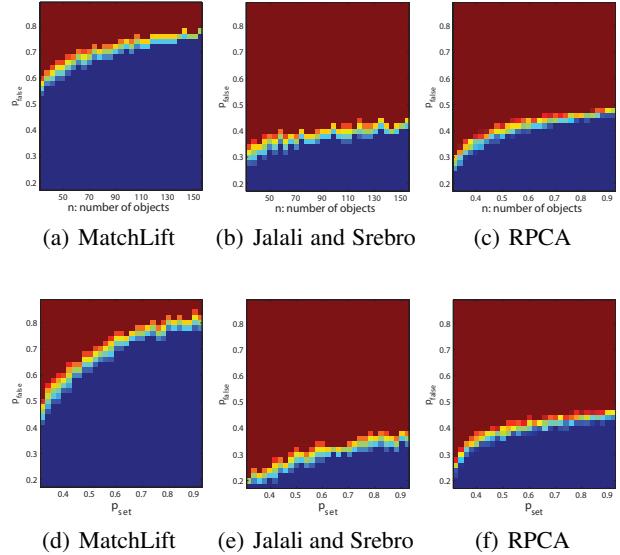


Figure 1. Phase Transition Diagrams of MatchLift, (Jalali & Srebro, 2012), and RPCA (Candès et al., 2011; Jalali et al., 2011). We can see that MatchLift can recover the ground-truth maps even when the majority of the input correspondences are wrong, while both (Jalali & Srebro, 2012) and RPCA require the input corruption rate to be less than 50%. (a-c)  $p_{\text{set}} = 0.6$ . (d-f)  $n = 100$ .

## 5. Experimental Evaluation

In this section, we evaluate the performance of MatchLift and compare it against (Candès et al., 2011; Jalali et al., 2011; Jalali & Srebro, 2012) and two other graph matching methods. We consider both synthetic examples, which are used to verify the exact recovery conditions described above, as well as popular benchmark datasets for evaluating the practicability on real-world images.

### 5.1. Synthetic Examples

We follow the randomized model described in Section 4 to generate synthetic examples. For simplicity, we only consider the full observation mode, which establishes input maps between all pairs of objects. In all examples, we fix the universe size such that it consists of  $m = 16$  points. We then vary the remaining parameters, i.e.,  $n$ ,  $p_{\text{set}}$  and  $p_{\text{false}}$ , to assess the performance of an algorithm. We evaluate  $31 \times 36$  sets of parameters for each scenario, where each parameter configuration is simulated by 10 Monte Carlo trials. The empirical success probability is reflected by the color of each cell. Blue denotes perfect recovery in all experiments, and red denotes failure for all trials.

Figure 1(a) illustrates the phase transition for  $p_{\text{set}} = 0.6$ , when the number of objects  $n$  and  $p_{\text{false}}$  vary. We can see that MatchLift is exact even when the majority of the input correspondences are incorrect (e.g., 75% when  $n = 150$ ). This is consistent with the theoretical result that the lower

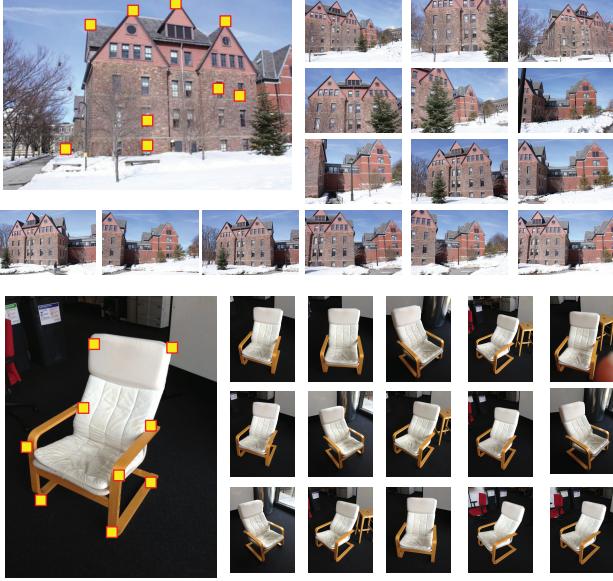


Figure 2. A small benchmark created for matching multiple images with partial similarity. Manually labeled feature points are highlighted. (Top) Building dataset (Bottom) Chair dataset.

bound on  $p_{\text{true}}$  for exact recovery is  $O(1/\sqrt{n})$ .

Figure 1(d) shows the phase transition, when  $p_{\text{set}}$  and  $p_{\text{false}}$  vary. We can see that MatchLift tolerates more noise when  $p_{\text{set}}$  is large. This is also consistent with the result that the error-correction ability of Matchlift improves with  $p_{\text{set}}$ .

Figures 1(b-c) and Figures 1(e-f) show the transition diagrams of (Jalali et al., 2011; Jalali & Srebro, 2012). One can see that MatchLift is empirically superior, as both (Jalali et al., 2011) and (Jalali & Srebro, 2012) do not allow dense error correction with respect to the particular noise model considered in this paper.

## 5.2. Real-World Examples

We have applied our algorithm on six benchmark datasets, i.e., CMU-House, CMU-Hotel, two datasets (Graf and Bikes) from (Mikolajczyk & Schmid, 2005) and two new datasets (referred as Chair and Building, respectively) designed for evaluating joint partial object matching. As shown in Figure 2, the first data set contains 16 images of a chair model from different viewpoints, and second data set contains 16 images taken around a building (Crandall et al., 2011). In the following, we first discuss the procedure for generating the input to our algorithm, i.e., the input sets and the initial maps. We then present the evaluation setup and analyze the results.

**Feature points and initial maps.** To make fair comparisons with previous techniques on CMU-House and CMU-Hotel, we use the features points provided in (Caetano et al., 2009) and apply the spectral matching algorithm de-

scribed in (Leordeanu & Hebert, 2005) to establish initial maps between features points. To assess the performance of the proposed algorithm with sparse input maps, we only match each image with 10 random neighboring images.

To handle raw images in Chair, Building, Graf and Bikes, we apply a different strategy to build feature points and initial maps. We first detect dense SIFT feature points (Lowe, 2004) on each image. We then apply RANSAC (Fischler & Bolles, 1981) to obtain correspondences between each pair of images. As SIFT feature points are over-complete, many of them do not appear in the resulting feature correspondences between pairs of views. Thus, we remove all feature points that have less than 2 appearances in all pairwise maps. We further apply farthest point sampling on the feature points until the sampling density is above  $0.05w$ , where  $w$  is the width of the input images. The remaining feature points turn out to be much more distinct and thus are suitable for joint matching (See Figure 3). For the experiments we have tested, we obtain about 60–100 features points per image.

**Evaluation protocol.** On CMU-House and CMU-Hotel, we count the percentage of correct feature correspondences produced by each algorithm. On Chair, Building, Graf and Bikes, we apply the metric described in (HaCohen et al., 2011), which evaluates the deviations of manual feature correspondences. As the feature points computed on each image do not necessarily align with the manual features, we apply (Ahmed et al., 2008) to interpolate feature level correspondences into pixel-wise correspondences for evaluation.

	Input	MatchLift	RPCA	LearnI	LearnII
House	68.2%	100%	92.2%	99.8%	96%
Hotel	64.1%	100%	90.1%	94.8%	90%

Table 1. Matching performance on the CMU-Hotel and CMU-House datasets. We compare the proposed MatchLift algorithm with RPCA and two learning based graph matching methods: LearnI (Leordeanu et al., 2012) and LearnII (Caetano et al., 2009).

**Results.** Table 1 shows the results of various algorithms on CMU-House and CMU-Hotel. We can see that even with moderate initial maps, MatchLift recovers all ground-truth correspondences. In contrast, the method of (Jalali et al., 2011) can only recover 92.2% and 90.1% ground-

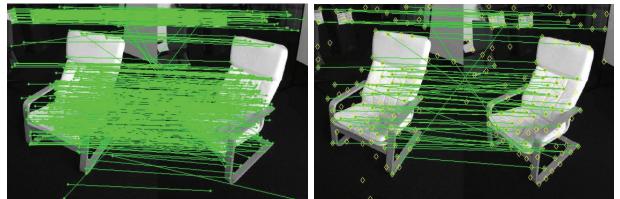


Figure 3. A map between dense SIFT feature points (Left) is converted into a map between sampled feature points (Right).

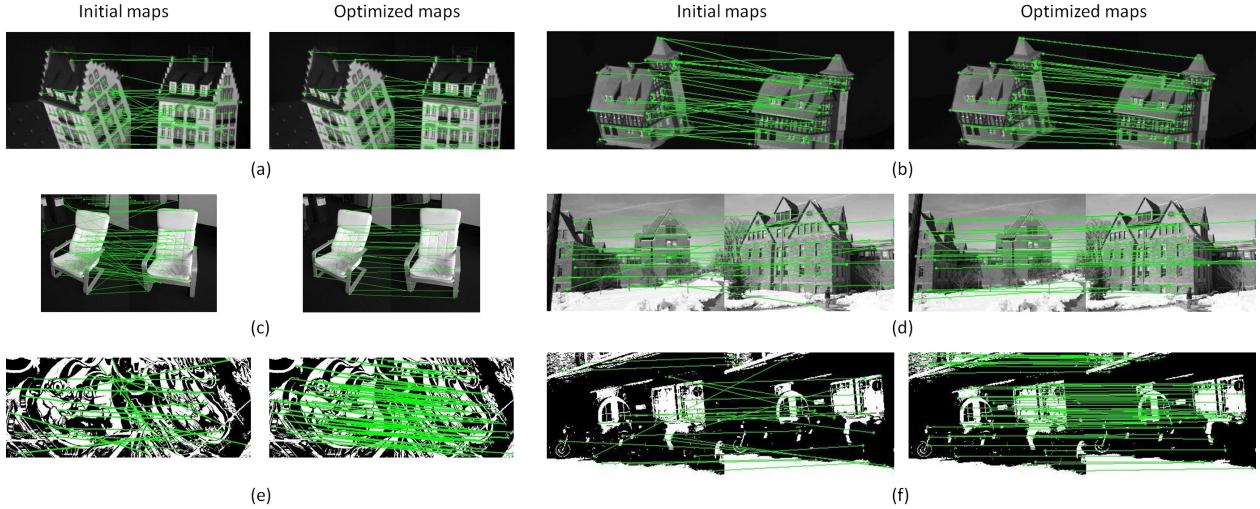


Figure 4. Comparisons between the input maps and the output of MatchLift on six benchmark datasets: (a) CMU Hotel, (b) CMU House, (c) Chair, (d) Building, (e) Graf, and (f) Bikes. The optimized maps not correct incorrect correspondences as well as fill in missing correspondences (generated by paths through intermediate shapes). One representative pair from each dataset is shown here.

truth correspondences on CMU-House and CMU-Hotel, respectively. Note that, MatchLift also outperforms state-of-the-art learning based graph matching algorithms (Cae-tano et al., 2009; Leordeanu et al., 2012). This shows the the advantage of the proposed approach.

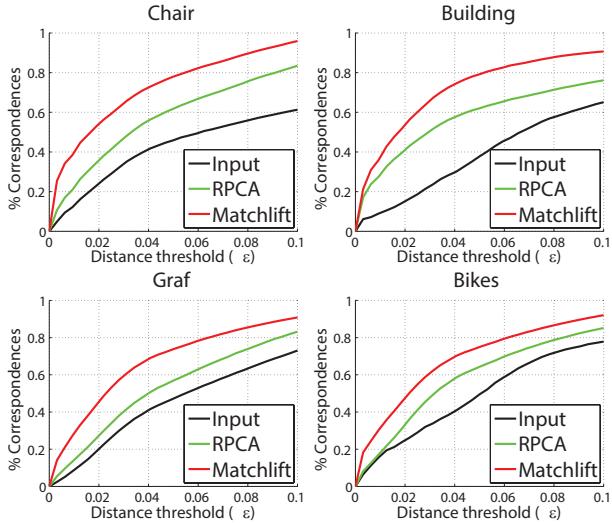


Figure 5. Percentages of ground-truth correspondences, whose distances to a map collection are below a varying threshold  $\epsilon$ .

Figure 4 and Figure 5 show the results of MatchLift on Chair, Building, Graf and Bikes. As these images contain noisy background information, the quality of the input maps is lower than those on House and Hotel. Encouragingly, MatchLift still recovers almost all manual correspondences. Moreover, MatchLift significantly outperforms (Jalali et al., 2011), as the fault-tolerance rate of

(Jalali et al., 2011) is limited by a small constant barrier.

Another interesting observation is that the improvements on Graf and Bikes (each has 6 images) are lower than those on Chair and Building (each has 16 images). This is consistent with the common knowledge of data-driven effect, where large object collections possess stronger self-correction power than small object collections.

## 6. Conclusions

This paper delivers some encouraging news: given a few noisy object matches computed in isolation, a collection of partially similar objects can be accurately matched via semidefinite relaxation — an approach which provably works under dense errors. The proposed algorithm is essentially parameter-free, and can be solved by ADMM while achieving remarkable efficiency and accuracy, with the assistance of a greedy rounding strategy.

The proposed algorithm achieves near-optimal error-correction ability, as it is guaranteed to work even when a dominant fraction of inputs are corrupted. This in turn underscores the importance of joint object matching: however low the quality of input sources is, perfect matching is achievable as long as we obtain sufficiently many instances. In a broader sense, our findings suggest that a large class of combinatorial and integer programming problems might be solved exactly and efficiently via semidefinite relaxation.

## Acknowledgements

This work was supported by ONR MURI N00014-13-1-0341, AFOSR FA9550-12-1-0372, NSF CCF 1161480 and DMS 1228304, Google and Motorola research awards.

## References

- Abbe, E., Bandeira, A., Singer, A., and Bracher, A. Linear inverse problems on Erdős-Rényi graphs: Information-theoretic limits and efficient recovery. *IEEE ISIT*, 2014.
- Ahmed, N., Theobalt, C., Rssl, C., Thrun, S., and Seidel, H. Dense correspondence finding for parametrization-free animation reconstruction from video. In *CVPR*, 2008.
- Bandeira, A., Charikar, M., Singer, A., and Zhu, A. Multireference alignment using semidefinite programming. In *arxiv:1308.5256*, 2013.
- Bansal, N., Blum, A., and Chawla, S. Correlation clustering. *Machine Learning*, 56(1-3):89–113, 2004.
- Caetano, T. S., McAuley, J., Cheng, L., Le, Q., and Smola, A. Learning graph matching. *PAMI*, 31(6), 2009.
- Candès, Emmanuel J., Li, Xiaodong, Ma, Yi, and Wright, John. Robust principal component analysis? *Journal of ACM*, 58(3):11:1–11:37, Jun 2011.
- Chaudhury, K., Khoo, Y., Singer, A., and Cowburn, D. Global registration of multiple point clouds using semidefinite programming. *arXiv:1306.5226*, 2013.
- Chen, Yudong, Sanghavi, Sujay, and Xu, Huan. Clustering sparse graphs. *NIPS*, 2012.
- Chen, Yudong, Jalali, Ali, Sanghavi, Sujay, and Caramanis, C. Low-rank matrix recovery from errors and erasures. *IEEE Transactions on Information Theory*, 59(7):4324–4337, July 2013.
- Chen, Yuxin and Goldsmith, Andrea J. Information recovery from pairwise measurements. *IEEE International Symposium on Information Theory (ISIT)*, July 2014.
- Chen, Yuxin, Guibas, Leonidas J., and Huang, Qixing. Near-optimal joint object matching via convex relaxation. *arxiv preprint arXiv:1402.1473*, 2014.
- Cho, T. S., Avidan, S., and Freeman, W. T. A probabilistic image jigsaw puzzle solver. *CVPR*, pp. 183–190, 2010.
- Cour, T., Srinivasan, P., and Shi, J. Balanced graph matching. *Neural Information Processing Systems*, 2007.
- Crandall, D., Owens, A., Snavely, N., and Huttenlocher, D. Discrete-continuous optimization for large-scale structure from motion. *IEEE CVPR*, pp. 3001–3008, 2011.
- Fischler, M. A. and Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.
- HaCohen, Y., Shechtman, E., Goldman, D., and Lischinski, D. Non-rigid dense correspondence with applications for image enhancement. *ACM Trans. Graph.*, 30(4):70:1–70:10, July 2011.
- Huang, Q. and Guibas, L. Consistent shape maps via semidefinite programming. *SGP*, 32(5):177–186, 2013.
- Huang, Q., Flöry, S., Gelfand, N., Hofer, M., and Pottmann, H. Reassembling fractured objects by geometric matching. *ACM Trans. Graph.*, 25(3), 2006.
- Huang, Q., Zhang, G., Gao, L., Hu, S.M., Butscher, A., and Guibas, L. An optimization approach for extracting and encoding consistent maps in a shape collection. *ACM Transactions on Graphics*, 31(6):167, 2012.
- Jalali, A., Chen, Y., Sanghavi, S., and Xu, H. Clustering partially observed graphs via convex optimization. *International Conf. on Machine Learning (ICML)*, 2011.
- Jalali, Ali and Srebro, Nati. Clustering using max-norm constrained optimization. *ICML*, June 2012.
- Keshavan, R. H., Montanari, A., and Oh, S. Matrix completion from noisy entries. *Journal of Machine Learning Research*, 99:2057–2078, 2010.
- Kim, V.G., Li, W., Mitra, N., DiVerdi, S., and Funkhouser, T. Exploring collections of 3d models using fuzzy correspondences. In *ACM SIGGRAPH*, 2012.
- Leordeanu, M. and Hebert, M. A spectral technique for correspondence problems using pairwise constraints. *ICCV*, pp. 1482–1489, 2005.
- Leordeanu, M., Sukthankar, R., and Hebert, M. Unsupervised learning for graph matching. *International journal of computer vision*, 96(1):28–45, 2012.
- Lowe, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.
- Mikolajczyk, K. and Schmid, C. A performance evaluation of local descriptors. *PAMI*, 27(10):1615–1630, 2005.
- Nguyen, A., Ben-Chen, M., Welniak, K., Ye, Y., and Guibas, L. An optimization approach to improving collections of shape maps. In *SGP*, pp. 1481–1491, 2011.
- Pachauri, D., Kondor, R., and Singh, V. Solving the multi-way matching problem by permutation synchronization. In *Neural Information Processing Systems (NIPS)*, 2013.
- Roberts, R., Sinha, S. N., Szeliski, R., and Steedly, D. Structure from motion for scenes with large duplicate structures. In *IEEE CVPR*, pp. 3137–3144, 2011.
- Wang, L. and Singer, A. Exact and stable recovery of rotations for robust synchronization. *arxiv:1211.2441*, 2013.
- Wen, Z., Goldfarb, D., and Yin, W. Alternating direction augmented Lagrangian methods for semidefinite programming. *Math. Prog. Comp.*, 2(3-4):203–230, 2010.
- Zach, C., Klopschitz, M., and Pollefeys, M. Disambiguating visual relations using loop constraints. In *IEEE CVPR*, pp. 1426–1433, 2010.