

# Biconvex Relaxation for Semidefinite Programming in Computer Vision

Sohil Shah<sup>\*1</sup>, Abhay Kumar Yadav<sup>\*1</sup>, Carlos D. Castillo<sup>1</sup>, David W. Jacobs<sup>1</sup>,  
Christoph Studer<sup>2</sup>, and Tom Goldstein<sup>1</sup>

<sup>1</sup> University of Maryland, College Park, MD USA; <sup>2</sup> Cornell University, Ithaca, NY USA.  
sohilas@umd.edu, {jaiabhay,djacobs,tomg}@cs.umd.edu,  
carlos@umiacs.umd.edu, studer@cornell.edu

**Abstract.** Semidefinite programming (SDP) is an indispensable tool in computer vision, but general-purpose solvers for SDPs are often too slow and memory intensive for large-scale problems. Our framework, referred to as biconvex relaxation (BCR), transforms an SDP consisting of PSD constraint matrices into a specific biconvex optimization problem, which can then be approximately solved in the original, low-dimensional variable space at low complexity. The resulting problem is solved using an efficient alternating minimization (AM) procedure. Since AM has the potential to get stuck in local minima, we propose a general initialization scheme that enables BCR to start close to a global optimum—this is key for BCR to quickly converge to optimal or near-optimal solutions. We showcase the efficacy of our approach on three applications in computer vision, namely segmentation, co-segmentation, and manifold metric learning. BCR achieves solution quality comparable to state-of-the-art SDP methods with speedups between  $4\times$  and  $35\times$ .

## 1 Introduction

Optimization problems involving either integer-valued vectors or low-rank matrices are ubiquitous in computer vision. Graph-cut methods for image segmentation, for example, involve optimization problems where integer-valued variables represent region labels [1,2,3,4]. Problems in multi-camera structure from motion [5], manifold embedding [6], and matrix completion [7] all rely on optimization problems involving matrices with low rank constraints. Since these constraints are non-convex, the design of efficient algorithms that find globally optimal solutions is a difficult task.

For a wide range of applications [8,9,10,11,12,6], non-convex constraints can be handled by *semidefinite relaxation* (SDR) [8]. In this approach, a non-convex optimization problem involving a vector of unknowns is “lifted” to a higher dimensional convex problem that involves a positive semidefinite (PSD) matrix, which then enables one to solve a SDP [13]. While SDR delivers state-of-the-art performance in a wide range of applications [4,8,14,3,6,7], the approach significantly increases the dimensionality of the original optimization problem (i.e., replacing a vector with a matrix), which typically results in exorbitant computational costs and memory requirements. Nevertheless, SDR leads to SDPs whose global optimal solution can be found using robust numerical methods.

---

\* The first two authors contributed equally to this work.

A growing number of computer-vision applications involve high-resolution images (or videos) that require SDPs with a large number of variables. General-purpose (interior point) solvers for SDPs do not scale well to such problem sizes; the worst-case complexity is  $O(N^{6.5} \log(1/\varepsilon))$  for an  $N \times N$  problem with  $\varepsilon$  objective error [15]. In imaging applications,  $N$  is often proportional to the number of pixels, which is potentially large.

The prohibitive complexity and memory requirements of solving SDPs exactly with a large number of variables has spawned interest in fast, non-convex solvers that avoid lifting. For example, recent progress in phase retrieval by Netrapalli *et al.* [16] and Candès *et al.* [17] has shown that non-convex optimization methods provably achieve solution quality comparable to exact SDR-based methods with significantly lower complexity. These methods operate on the original dimensions of the (un-lifted) problem, which enables their use on high-dimensional problems. Another prominent example is max-norm regularization by Lee *et al.* [18], which was proposed for solving high-dimensional matrix-completion problems and to approximately perform max-cut clustering. This method was shown to outperform exact SDR-based methods in terms of computational complexity, while delivering acceptable solution quality. While both of these examples outperform classical SDP-based methods, they are limited to very specific problem types, and cannot handle more complex SDPs that typically appear in computer vision.

## 1.1 Contributions

We introduce a novel framework for approximately solving SDPs with positive semi-definite constraint matrices in a computationally efficient manner and with small memory footprint. Our proposed *bi-convex* relaxation (BCR), transforms an SDP into a biconvex optimization problem, which can then be solved in the original, low-dimensional variable space at low complexity. The resulting biconvex problem is solved using a computationally-efficient AM procedure. Since AM is prone to get stuck in local minima, we propose an initialization scheme that enables BCR to start close to the global optimum of the original SDP—this initialization is key for our algorithm to quickly converge to an optimal or near-optimal solution. We showcase the effectiveness of the BCR framework by comparing to highly-specialized SDP solvers for a selected set of problems in computer vision involving image segmentation, co-segmentation, and metric learning on manifolds. Our results demonstrate that BCR enables high-quality results while achieving speedups ranging from  $4\times$  to  $35\times$  over state-of-the-art competitor methods [19,20,21,22,23] for the studied applications.

## 2 Background and Relevant Prior Art

We now briefly review semidefinite programs (SDPs) and discuss prior work on fast, approximate solvers for SDPs in computer vision and related applications.

### 2.1 Semidefinite Programs (SDPs)

SDPs find use in a large and growing number of fields, including computer vision, machine learning, signal and image processing, statistics, communications, and control [13].

SDPs can be written in the following general form:

$$\begin{aligned} & \underset{\mathbf{Y} \in \mathcal{S}_{N \times N}^+}{\text{minimize}} && \langle \mathbf{C}, \mathbf{Y} \rangle \\ & \text{subject to} && \langle \mathbf{A}_i, \mathbf{Y} \rangle = b_i, \quad \forall i \in \mathcal{E}, \\ & && \langle \mathbf{A}_j, \mathbf{Y} \rangle \leq b_j, \quad \forall j \in \mathcal{B}, \end{aligned} \tag{1}$$

where  $\mathcal{S}_{N \times N}^+$  represents the set of  $N \times N$  symmetric positive semidefinite matrices, and  $\langle \mathbf{C}, \mathbf{Y} \rangle = \text{tr}(\mathbf{C}^T \mathbf{Y})$  is the matrix inner product. The sets  $\mathcal{E}$  and  $\mathcal{B}$  contain the indices associated with the equality and inequality constraints, respectively;  $\mathbf{A}_i$  and  $\mathbf{A}_j$  are symmetric matrices of appropriate dimensions.

The key advantages of SDPs are that (i) they enable the transformation of certain non-convex constraints into convex constraints via semidefinite relaxation (SDR) [8] and (ii) the resulting problems often come with strong theoretical guarantees.

In computer vision, a large number of problems can be cast as SDPs of the general form (1). For example, [6] formulates image manifold learning as an SDP, [12] uses an SDP to enforce a non-negative lighting constraint when recovering scene lighting and object albedos, [24] uses an SDP for graph matching, [5] proposes an SDP that recovers the orientation of multiple cameras from point correspondences and essential matrices, and [7] uses low-rank SDPs to solve matrix-completion problems that arise in structure-from-motion and photometric stereo.

## 2.2 SDR for Binary-Valued Quadratic Problems

Semidefinite relaxation is commonly used to solve binary-valued labeling problems. For such problems, a set of variables take on binary values while minimizing a quadratic cost function that depends on the assignment of pairs of variables. Such labeling problems typically arise from Markov random fields (MRFs) for which many solution methods exist [25]. Spectral methods, e.g., [1], are often used to solve such binary-valued quadratic problems (BQPs)—the references [2,3] used SDR inspired by the work of [4] that provides a generalized SDR for the max-cut problem. BQP problems have wide applicability to computer vision problems, such as segmentation and perceptual organization [2,19,26], semantic segmentation [27], matching [3,28], surface reconstruction including photometric stereo and shape from defocus [11], and image restoration [29].

BQPs can be solved by lifting the binary-valued label vector  $\mathbf{b} \in \{\pm 1\}^N$  to an  $N^2$ -dimensional matrix space by forming the PSD matrix  $\mathbf{B} = \mathbf{b}\mathbf{b}^T$ , whose non-convex rank-1 constraint is relaxed to PSD matrices  $\mathbf{B} \in \mathcal{S}_{N \times N}^+$  with an all-ones diagonal [8]. The goal is then to solve a SDP for  $\mathbf{B}$  in the hope that the resulting matrix has rank 1; if  $\mathbf{B}$  has higher rank, an approximate solution must be extracted which can either be obtained from the leading eigenvector or via randomization methods [8,30].

## 2.3 Specialized Solvers for SDPs

General-purpose solvers for SDPs, such as SeDuMi [31] or SDPT3 [32], rely on interior point methods with high computational complexity and memory requirements. Hence, their use is restricted to low-dimensional problems. For problems in computer vision,

where the number of variables can become comparable to the number of pixels in an image, more efficient algorithms are necessary. A handful of special-purpose algorithms have been proposed to solve specific problem types arising in computer vision. These algorithms fit into two classes: (i) convex algorithms that solve the original SDP by exploiting problem structure and (ii) non-convex methods that avoid lifting.

For certain problems, one can exactly solve SDPs with much lower complexity than interior point schemes, especially for BQP problems in computer vision. Ecker *et al.* [11] deployed a number of heuristics to speed up the Goemans-Williamson SDR [4] for surface reconstruction. Olsson *et al.* [29] proposed a spectral subgradient method to solve BQP problems that include a linear term, but are unable to handle inequality constraints. A particularly popular approach is the SDCut algorithms of Wang *et al.* [19]. This method solves BQP for some types of segmentation problems using dual gradient descent. SDCut leads to a similar relaxation as for BQP problems, but enables significantly lower complexity for graph cutting and its variants. To the best of our knowledge, the method by Wang *et al.* [19] yields state-of-the-art performance—nevertheless, our proposed method is at least an order of magnitude faster, as shown in Section 4.

Another algorithm class contains non-convex approximation methods that avoid lifting altogether. Since these methods work with low-dimensional unknowns, they are potentially more efficient than lifted methods. Simple examples include the Wiberg method [33] for low-rank matrix approximation, which uses Newton-type iterations to minimize a non-convex objective. A number of methods have been proposed for SDPs where the objective function is simply the trace-norm of  $\mathbf{Y}$  (i.e., problem (1) with  $\mathbf{C} = \mathbf{I}$ ) and without inequality constraints. Approaches include replacing the trace norm with the max-norm [18], or using the so-called Wirtinger flow to solve phase-retrieval problems [17]. One of the earliest approaches for non-convex methods are due to Burer and Montiero [34], who propose an augmented Lagrangian method. While this method is able to handle arbitrary objective functions, it does not naturally support inequality constraints (without introducing auxiliary slack variables). Furthermore, this approach uses convex methods for which convergence is not well understood and is sensitive to the initialization value.

While most of the above-mentioned methods provide best-in-class performance at low computational complexity, they are limited to very specific problems and cannot be generalized to other, more general SDPs.

### 3 Biconvex Relaxation (BCR) Framework

We now present the proposed *biconvex relaxation (BCR)* framework. We then propose an alternating minimization procedure and a suitable initialization method.

#### 3.1 Biconvex Relaxation

Rather than solving the general SDP (1) directly, we exploit the following key fact: any matrix  $\mathbf{Y}$  is symmetric positive semidefinite if and only if it has an expansion of the form  $\mathbf{Y} = \mathbf{X}\mathbf{X}^T$ . By substituting the factorization  $\mathbf{Y} = \mathbf{X}\mathbf{X}^T$  into (1), we are able to

remove the semidefinite constraint and arrive at the following problem:

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{N \times r}}{\text{minimize}} && \text{tr}(\mathbf{X}^T \mathbf{C} \mathbf{X}) \\ & \text{subject to} && \text{tr}(\mathbf{X}^T \mathbf{A}_i \mathbf{X}) = b_i, \quad \forall i \in \mathcal{E}, \\ & && \text{tr}(\mathbf{X}^T \mathbf{A}_j \mathbf{X}) \leq b_j, \quad \forall j \in \mathcal{B}, \end{aligned} \quad (2)$$

where  $r = \text{rank}(\mathbf{Y})$ .<sup>1</sup> Note that any symmetric semi-definite matrix  $\mathbf{A}$  has a (possibly complex-valued) square root  $\mathbf{L}$  of the form  $\mathbf{A} = \mathbf{L}^T \mathbf{L}$ . Furthermore, we have  $\text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{X}^T \mathbf{L}^T \mathbf{L} \mathbf{X}) = \|\mathbf{L} \mathbf{X}\|_F^2$ , where  $\|\cdot\|_F$  is the Frobenius (matrix) norm. This formulation enables us to rewrite (2) as follows:

$$\begin{aligned} & \underset{\mathbf{X} \in \mathbb{R}^{N \times r}}{\text{minimize}} && \text{tr}(\mathbf{X}^T \mathbf{C} \mathbf{X}) \\ & \text{subject to} && \mathbf{Q}_i = \mathbf{L}_i \mathbf{X}, \quad \|\mathbf{Q}_i\|_F^2 = b_i, \quad \forall i \in \mathcal{E}, \\ & && \mathbf{Q}_j = \mathbf{L}_j \mathbf{X}, \quad \|\mathbf{Q}_j\|_F^2 \leq b_j, \quad \forall j \in \mathcal{B}. \end{aligned} \quad (3)$$

If the matrices  $\{\mathbf{A}_i\}$ ,  $\{\mathbf{A}_j\}$ , and  $\mathbf{C}$  are themselves PSDs, then the objective function in (3) is convex and quadratic, and the inequality constraints in (3) are convex—non-convexity of the problem is only caused by the equality constraints. The core idea of BCR explained next is to relax these equality constraints. Here, we assume that the factors of these matrices are easily obtained from the underlying problem structure. For some applications, where these factors are not readily available this could be a computational burden (worst case  $\mathcal{O}(N^3)$ ) rather than an asset.

In the formulation (3), we have lost convexity. Nevertheless, whenever  $r < N$ , we achieved a (potentially large) dimensionality reduction compared to the original SDP (1). We now relax (3) in a form that is biconvex, i.e., convex with respect to a group of variables when the remaining variables are held constant. By relaxing the convex problem in biconvex form, we retain many advantages of the convex formulation while maintaining low dimensionality and speed. In particular, we propose to approximate (3) with the following *biconvex relaxation (BCR)*:

$$\begin{aligned} & \underset{\mathbf{x}, \mathbf{Q}_i, i \in \{\mathcal{B} \cup \mathcal{E}\}}{\text{minimize}} && \text{tr}(\mathbf{X}^T \mathbf{C} \mathbf{X}) + \frac{\alpha}{2} \sum_{i \in \{\mathcal{E} \cup \mathcal{B}\}} \|\mathbf{Q}_i - \mathbf{L}_i \mathbf{X}\|_F^2 - \frac{\beta}{2} \sum_{j \in \mathcal{E}} \|\mathbf{Q}_j\|_F^2 \\ & \text{subject to} && \|\mathbf{Q}_i\|_F^2 \leq b_i, \quad \forall i \in \{\mathcal{B} \cup \mathcal{E}\}, \end{aligned} \quad (4)$$

where  $\alpha > \beta > 0$  are relaxation parameters (discussed in detail below). In this BCR formulation, we relaxed the equality constraints  $\|\mathbf{Q}_i\|_F^2 = b_i, \forall i \in \mathcal{E}$ , to inequality constraints  $\|\mathbf{Q}_i\|_F^2 \leq b_i, \forall i \in \mathcal{E}$ , and added negative quadratic penalty functions  $-\frac{\beta}{2} \|\mathbf{Q}_j\|_F^2, \forall j \in \mathcal{E}$ , to the objective function. These quadratic penalties attempt to force the inequality constraints in  $\mathcal{E}$  to be satisfied exactly. We also replaced the constraints  $\mathbf{Q}_i = \mathbf{L}_i \mathbf{X}$  and  $\mathbf{Q}_j = \mathbf{L}_j \mathbf{X}$  by quadratic penalty functions in the objective function.

The relaxation parameters are chosen by freezing the ratio  $\alpha/\beta$  to 2, and following a simple, principled way of setting  $\beta$ . Unless stated otherwise, we set  $\beta$  to match the

<sup>1</sup> Straightforward extensions of our approach allow us to handle constraints of the form  $\text{tr}(\mathbf{X}^T \mathbf{A}_k \mathbf{X}) \geq b_k, \forall k \in \mathcal{A}$ , as well as complex-valued matrices and vectors.

curvature of the penalty term with the curvature of the objective i.e.,  $\beta = \|\mathbf{C}\|_2$ , so that the resulting bi-convex problem is well-conditioned.

Our BCR formulation (4) has some important properties. First, if  $\mathbf{C} \in \mathcal{S}_{N \times N}^+$  then the problem is biconvex, i.e., convex with respect to  $\mathbf{X}$  when the  $\{\mathbf{Q}_i\}$  are held constant, and vice versa. Furthermore, consider the case of solving a constraint feasibility problem (i.e., problem (1) with  $\mathbf{C} = \mathbf{0}$ ). When  $\mathbf{Y} = \mathbf{X}\mathbf{X}^T$  is a solution to (1) with  $\mathbf{C} = \mathbf{0}$ , the problem (4) assumes objective value  $-\frac{\beta}{2} \sum_j b_j$ , which is the global minimizer of the BCR formulation (4). Likewise, it is easy to see that any global minimizer of (4) with objective value  $-\frac{\beta}{2} \sum_j b_j$  must be a solution to the original problem (1).

### 3.2 Alternating Minimization (AM) Algorithm

One of the key benefits of biconvexity is that (4) can be globally minimized with respect to  $\mathbf{Q}$  or  $\mathbf{X}$ . Hence, it is natural to compute approximate solutions to (4) via alternating minimization. Note the convergence of AM for biconvex problems is well understood [35,36]. The two stages of the proposed method for BCR are detailed next.

**Stage 1: Minimize with respect to  $\{\mathbf{Q}_i\}$ .** The BCR objective in (4) is quadratic in  $\{\mathbf{Q}_i\}$  with no dependence between matrices. Consequently, the optimal value of  $\mathbf{Q}_i$  can be found by minimizing the quadratic objective, and then reprojecting back into a unit Frobenius-norm ball of radius  $\sqrt{b_i}$ . The minimizer of the quadratic objective is given by  $\frac{\alpha}{\alpha - \beta_i} \mathbf{L}_i \mathbf{X}$ , where  $\beta_i = 0$  if  $i \in \mathcal{B}$  and  $\beta_i = \beta$  if  $i \in \mathcal{E}$ . The projection onto the unit ball then leads to the following *expansion–reprojection* update:

$$\mathbf{Q}_i \leftarrow \frac{\mathbf{L}_i \mathbf{X}}{\|\mathbf{L}_i \mathbf{X}\|_F} \min \left\{ \sqrt{b_i}, \frac{\alpha}{\alpha - \beta_i} \|\mathbf{L}_i \mathbf{X}\|_F \right\}. \quad (5)$$

Intuitively, this expansion–reprojection update causes the matrix  $\mathbf{Q}_i$  to expand if  $i \in \mathcal{E}$ , thus encouraging it to satisfy the relaxed constraints in (4) with equality.

**Stage 2: Minimize with respect to  $\mathbf{X}$ .** This stage solves the least-squares problem:

$$\mathbf{X} \leftarrow \arg \min_{\mathbf{X} \in \mathbb{R}^{N \times r}} \text{tr}(\mathbf{X}^T \mathbf{C} \mathbf{X}) + \frac{\alpha}{2} \sum_{i \in \{\mathcal{E} \cup \mathcal{B}\}} \|\mathbf{Q}_i - \mathbf{L}_i \mathbf{X}\|_F^2. \quad (6)$$

The optimality conditions for this problem are linear equations, and the solution is

$$\mathbf{X} \leftarrow \left( \mathbf{C} + \alpha \sum_{i \in \{\mathcal{E} \cup \mathcal{B}\}} \mathbf{L}_i^T \mathbf{L}_i \right)^{-1} \left( \sum_{i \in \{\mathcal{E} \cup \mathcal{B}\}} \mathbf{L}_i^T \mathbf{Q}_i \right), \quad (7)$$

where the matrix inverse (one-time computation) may be replaced by a pseudo-inverse if necessary. Alternatively, one may perform a simple gradient-descent step with a suitable step size, which avoids the inversion of a potentially large-dimensional matrix.

The resulting AM algorithm for the proposed BCR (4) is summarized in Algorithm 1.

### 3.3 Initialization

The problem (4) is biconvex and hence, a global minimizer can be found with respect to either  $\{\mathbf{Q}_i\}$  or  $\mathbf{X}$ , although a global minimizer of the joint problem is not guaranteed.

**Algorithm 1** AM for Biconvex Relaxation

- 
- 1: **inputs:**  $\mathbf{C}$ ,  $\{\mathbf{L}_i\}$ ,  $b_i$ ,  $\alpha$ , and  $\beta$ , **output:**  $\mathbf{X}$
  - 2: Compute an initializer for  $\mathbf{X}$  as in Section 3.3
  - 3: Precompute  $\mathbf{M} = (\mathbf{C} + \alpha \sum_{i \in \{E \cup B\}} \mathbf{L}_i^T \mathbf{L}_i)^{-1}$
  - 4: **while** not converged **do**
  - 5:    $\mathbf{Q}_i \leftarrow \frac{\mathbf{L}_i \mathbf{X}}{\|\mathbf{L}_i \mathbf{X}\|_F} \min \left\{ \sqrt{b_i}, \frac{\alpha}{\alpha - \beta_i} \|\mathbf{L}_i \mathbf{X}\|_F \right\}$
  - 6:    $\mathbf{X} \leftarrow \mathbf{M} \left( \sum_{i \in \{E \cup B\}} \mathbf{L}_i^T \mathbf{Q}_i \right)$ ,
  - 7: **end while**
- 

We hope to find a global minimizer at low complexity using the AM method, but in practice AM may get trapped in local minima, especially if the variables have been initialized poorly. We now propose a principled method for computing an initializer for  $\mathbf{X}$  that is often close to the global optimum of the BCR problem—our initializer is key for the success of the proposed AM procedure and enables fast convergence.

The papers [16,17] have considered optimization problems that arise in phase retrieval where  $\mathcal{B} = \emptyset$  (i.e., there are only equality constraints),  $\mathbf{C} = \mathbf{I}$  being the identity, and  $\mathbf{Y}$  being rank one. For such problems, the objective of (1) reduces to  $\text{tr}(\mathbf{Y})$ . By setting  $\mathbf{Y} = \mathbf{x}\mathbf{x}^T$ , we obtain the following formulation:

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimize}} \|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{q}_i = \mathbf{L}_i \mathbf{x}, \quad \|\mathbf{q}_i\|_2^2 = b_i, \quad \forall i \in \mathcal{E}. \quad (8)$$

Netrapali *et al.* [16] proposed an iterative algorithm for solving (8), which has been initialized by the following strategy. Define

$$\mathbf{Z} = \frac{1}{|\mathcal{E}|} \sum_{i \in \mathcal{E}} b_i \mathbf{L}_i^T \mathbf{L}_i. \quad (9)$$

Let  $\mathbf{v}$  be the leading eigenvector of  $\mathbf{Z}$  and  $\lambda$  the leading eigenvalue. Then  $\mathbf{x} = \lambda \mathbf{v}$  is an accurate approximation to the true solution of (8). In fact, if the matrices  $\mathbf{L}_i$  are sampled from a random normal distribution, then it was shown in [16,17] that  $\mathbb{E} \|\mathbf{x}^* - \lambda \mathbf{x}\|_2^2 \rightarrow 0$  (in expectation) as  $|\mathcal{E}| \rightarrow \infty$ , where  $\mathbf{x}^*$  is the true solution to (8).

We are interested in a good initializer for the general problem in (3) where  $\mathbf{X}$  can be rank one or higher. We focus on problems with equality constraints only—note that one can use slack variables to convert a problem with inequality constraints into the same form [13]. Given that  $\mathbf{C}$  is a symmetric positive definite matrix, it can be decomposed into  $\mathbf{C} = \mathbf{U}^T \mathbf{U}$ . By the change of variables  $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}$ , we can rewrite (1) as follows:

$$\underset{\mathbf{X} \in \mathbb{R}^{N \times r}}{\text{minimize}} \|\tilde{\mathbf{X}}\|_F^2 \quad \text{subject to} \quad \langle \tilde{\mathbf{A}}_i, \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \rangle = b_i, \quad \forall i \in \mathcal{E}, \quad (10)$$

where  $\tilde{\mathbf{A}}_i = \mathbf{U}^{-T} \mathbf{A}_i \mathbf{U}^{-1}$ , and we omitted the inequality constraints. To initialize the proposed AM procedure in Algorithm 1, we make the change of variables  $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}$  to transform the BCR formulation into the form of (10). Analogously to the initialization procedure in [16] for phase retrieval, we then compute an initializer  $\tilde{\mathbf{X}}_0$  using the leading  $r$  eigenvectors of  $\mathbf{Z}$  scaled by the leading eigenvalue  $\lambda$ . Finally, we calculate the initializer for the original problem by reversing the change of variables as  $\mathbf{X}_0 = \mathbf{U}^{-1} \tilde{\mathbf{X}}_0$ . For most problems the initialization time is a small fraction of the total runtime.

### 3.4 Advantages of Biconvex Relaxation

The proposed framework has numerous advantages over other non-convex methods. First and foremost, BCR can be applied to general SDPs. Specialized methods, such as Wirtinger flow [17] for phase retrieval and the Wiberg method [33] for low-rank approximation are computationally efficient, but restricted to specific problem types. Similarly, the max-norm method [18] is limited to solving trace-norm-regularized SDPs. The method of Burer and Montiero [34] is less specialized, but does not naturally support inequality constraints. Furthermore, since BCR problems are biconvex, one can use numerical solvers with guaranteed convergence. Convergence is guaranteed not only for the proposed AM least-squares method in Algorithm 1 (for which the objective decreases monotonically), but also for a broad range of gradient-descent schemes suitable to find solutions to biconvex problems [37]. In contrast, the method in [34] uses augmented Lagrangian methods with non-linear constraints for which convergence is not guaranteed.

## 4 Benchmark Problems

We now evaluate our solver using both synthetic and real-world data. We begin with a brief comparison showing that biconvex solvers outperform both interior-point methods for general SDPs and also state-of-the-art low-rank solvers. Of course, specialized solvers for specific problem forms achieve superior performance to classical interior point schemes. For this reason, we evaluate our proposed method on three important computer vision applications, i.e., segmentation, co-segmentation, and manifold metric learning, using public datasets, and we compare our results to state-of-the-art methods. These applications are ideal because (i) they involve large scale SDPs and (ii) customized solvers are available that exploit problem structure to solve these problems efficiently. Hence, we can compare our BCR framework to powerful and optimized solvers.

### 4.1 General-Form Problems

We briefly demonstrate that BCR performs well on general SDPs by comparing to the widely used SDP solver, SDPT3 [32] and the state-of-the-art, low-rank SDP solver CGDSP [38]. Note that SDPT3 uses an interior point approach to solve the convex problem in (1) whereas the CGDSP solver uses gradient-descent to solve a non-convex formulation. For fairness, we initialize both algorithms using the proposed initializer and the gradient descent step in CGDSP was implemented using various acceleration techniques [39]. Since CGDSP cannot handle inequality constraints we restrict our comparison to equality constraints only.

**Experiments:** We randomly generate a  $256 \times 256$  rank-3 data matrix of the form  $\mathbf{Y}_{\text{true}} = \mathbf{x}_1 \mathbf{x}_1^T + \mathbf{x}_2 \mathbf{x}_2^T + \mathbf{x}_3 \mathbf{x}_3^T$ , where  $\{\mathbf{x}_i\}$  are standard normal vectors. We generate a standard normal matrix  $\mathbf{L}$  and compute  $\mathbf{C} = \mathbf{L}^T \mathbf{L}$ . Gaussian matrices  $\mathbf{A}_i \in \mathbb{R}^{250 \times 250}$  form equality constraints. We report the relative error in the recovered solution  $\mathbf{Y}_{\text{rec}}$  measured as  $\|\mathbf{Y}_{\text{rec}} - \mathbf{Y}_{\text{true}}\| / \|\mathbf{Y}_{\text{true}}\|$ . Average runtimes for varying numbers of constraints are shown in Figure 1a, while Figure 1b plots the average relative error. Figure 1a shows that our method has the best runtime of all the schemes. Figure 1b shows convex interior



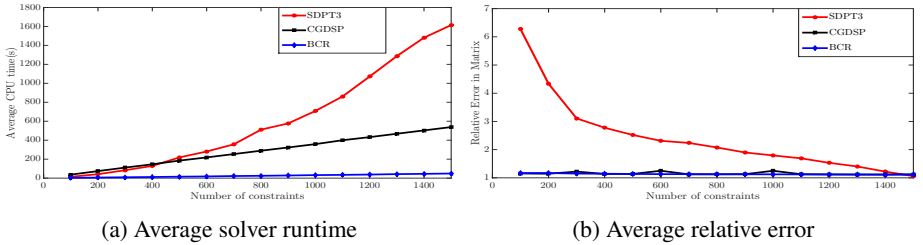


Fig. 1: Results on synthetic data for varying number of linear constraints.

point methods do not recover the correct solution for small numbers of constraints. With few constraints, the full lifted SDP is under-determined, allowing the objective to go to zero. In contrast, the proposed BCR approach is able to enforce an additional rank-3 constraint, which is advantageous when the number of constraints is low.

## 4.2 Image Segmentation

Consider an image of  $N$  pixels. Segmentation of foreground and background objects can be accomplished using graph-based approaches, where graph edges encode the similarities between pixel pairs. Such approaches include normalized cut [1] and ratio cut [40]. The graph cut problem can be formulated as an NP-hard integer program [4]

$$\begin{aligned} & \text{minimize } \mathbf{x}^T \mathbf{L} \mathbf{x}, \\ & \mathbf{x} \in \{-1, 1\}^N \end{aligned} \quad (11)$$

where  $\mathbf{L}$  encodes edge weights and  $\mathbf{x}$  contains binary region labels, one for each pixel. This problem can be “lifted” to the equivalent higher dimensional problem

$$\begin{aligned} & \text{minimize } \text{tr}(\mathbf{L}^T \mathbf{X}) \quad \text{subject to } \text{diag}(\mathbf{X}) = \mathbf{1}, \quad \text{rank}(\mathbf{X}) = 1. \\ & \mathbf{X} \in \mathcal{S}_{N \times N}^+ \end{aligned} \quad (12)$$

After dropping the non-convex rank constraint, (12) becomes an SDP that is solvable using convex optimization [14, 2, 28]. The SDP approach is computationally intractable if solved using off-the-shelf SDP solvers (such as SDPT3 [32] or other interior point methods). Furthermore, exact solutions cannot be recovered when the solution to the SDP has rank greater than 1. In contrast, BCR is computational efficient for large problems and can easily incorporate rank constraints, leading to efficient spectral clustering.

BCR is also capable of incorporating annotated foreground and background pixel priors [41] using linear equality and inequality constraints. We consider the SDP based segmentation presented in [41], which contains three grouping constraints on the pixels:  $(\mathbf{t}_f^T \mathbf{P} \mathbf{x})^2 \geq \kappa \|\mathbf{t}_f^T \mathbf{P} \mathbf{x}\|_1^2$ ,  $(\mathbf{t}_b^T \mathbf{P} \mathbf{x})^2 \geq \kappa \|\mathbf{t}_b^T \mathbf{P} \mathbf{x}\|_1^2$  and  $((\mathbf{t}_f - \mathbf{t}_b)^T \mathbf{P} \mathbf{x})^2 \geq \kappa \|(\mathbf{t}_f - \mathbf{t}_b)^T \mathbf{P} \mathbf{x}\|_1^2$ , where  $\kappa \in [0, 1]$ .  $\mathbf{P} = \mathbf{D}^{-1} \mathbf{W}$  is the normalized pairwise affinity matrix and  $\mathbf{t}_f$  and  $\mathbf{t}_b$  are indicator variables denoting the foreground and background pixels. These constraints enforce that the segmentation respects the pre-labeled pixels given by the user, and also pushes high similarity pixels to have the same label. The affinity matrix  $\mathbf{W}$  is given by

$$W_{i,j} = \begin{cases} \exp\left(-\frac{\|\mathbf{f}_i - \mathbf{f}_j\|_2^2}{\gamma_f^2} - \frac{d(i,j)^2}{\gamma_d^2}\right), & \text{if } d(i,j) < r \\ 0, & \text{otherwise,} \end{cases} \quad (13)$$

where  $\mathbf{f}_i$  is the color histogram of the  $i$ th super-pixel and  $d(i, j)$  is the spatial distance between  $i$  and  $j$ . Considering these constraints and letting  $\mathbf{X} = \mathbf{Y}\mathbf{Y}^T$ , (12) can be written in the form of (2) as follows:

$$\begin{aligned} & \underset{\mathbf{Y} \in \mathbb{R}^{N \times r}}{\text{minimize}} && \text{tr}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) \\ & \text{subject to} && \text{tr}(\mathbf{Y}^T \mathbf{A}_i \mathbf{Y}) = 1, \quad \forall i = 1, \dots, N \\ & && \text{tr}(\mathbf{Y}^T \mathbf{B}_2 \mathbf{Y}) \geq \kappa \|\mathbf{t}_f^T \mathbf{P} \mathbf{x}\|_1^2, \quad \text{tr}(\mathbf{Y}^T \mathbf{B}_3 \mathbf{Y}) \geq \kappa \|\mathbf{t}_b^T \mathbf{P} \mathbf{x}\|_1^2 \\ & && \text{tr}(\mathbf{Y}^T \mathbf{B}_4 \mathbf{Y}) \geq \kappa \|(\mathbf{t}_f - \mathbf{t}_b)^T \mathbf{P} \mathbf{x}\|_1^2, \quad \text{tr}(\mathbf{Y}^T \mathbf{B}_1 \mathbf{Y}) = 0. \end{aligned} \quad (14)$$

Here,  $r$  is the rank of the desired solution,  $\mathbf{B}_1 = \mathbf{1}\mathbf{1}^T$ ,  $\mathbf{B}_2 = \mathbf{P}\mathbf{t}_f\mathbf{t}_f^T\mathbf{P}$ ,  $\mathbf{B}_3 = \mathbf{P}\mathbf{t}_b\mathbf{t}_b^T\mathbf{P}$ ,  $\mathbf{B}_4 = \mathbf{P}(\mathbf{t}_f - \mathbf{t}_b)(\mathbf{t}_f - \mathbf{t}_b)^T\mathbf{P}$ ,  $\mathbf{A}_i = \mathbf{e}_i\mathbf{e}_i^T$ ,  $\mathbf{e}_i \in \mathbb{R}^n$  is an elementary vector with a 1 at the  $i$ th position. After solving (14) using BCR (4), the final binary solution is extracted from the score vector using the swept random hyperplanes method [30].

We compare the performance of BCR with the highly customized BQP solver SDCut [19] and biased normalized cut (BNCut) [20]. BNCut is an extension of the Normalized cut algorithm [1] whereas SDCut is currently the most efficient and accurate SDR solver but limited only to solving BQP problems. Also, BNCut can support only one quadratic grouping constraint per problem.

**Experiments:** We consider the Berkeley image segmentation dataset [42]. Each image is segmented into super-pixels using the VL-Feat [43] toolbox. For SDCut and BNCut, we use the publicly available code with hyper-parameters set to the values suggested in [19]. For BCR, we set  $\beta = \lambda/\sqrt{|\mathcal{B} \cup \mathcal{E}|}$ , where  $\lambda$  controls the coarseness of the segmentation by mediating the tradeoff between the objective and constraints, and would typically be chosen from [1, 10] via cross validation. For simplicity, we just set  $\lambda = 5$  in all experiments reported here.

We compare the runtime and quality of each algorithm. Figure 2 shows the segmentation results while the quantitative results are displayed in Table 1. For all the considered images, our approach gives superior foreground object segmentation compared to SDCut and BNCut. Moreover, as seen in Table 1, our solver is  $35\times$  faster than SDCut and yields lower objective energy. Segmentation using BCR is achieved using only rank 2 solutions whereas SDCut requires rank 7 solutions to obtain results of comparable accuracy.<sup>2</sup> Note that while BNCut with rank 1 solutions is much faster than SDP based methods, the BNCut segmentation results are not on par with SDP approaches.

### 4.3 Co-segmentation

We next consider image co-segmentation, in which segmentation of the same object is jointly computed on multiple images simultaneously. Because co-segmentation involves multiple images, it provides a testbed for large problem instances. Co-segmentation balances a tradeoff between two criteria: (i) color and spatial consistency within a single image and (ii) discrimination between foreground and background pixels over multiple images. We closely follow the work of Joulin *et al.* [26], whose formulation is given by

$$\underset{\mathbf{x} \in \{\pm 1\}^N}{\text{minimize}} \quad \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \text{subject to} \quad (\mathbf{x}^T \delta_i)^2 \leq \lambda^2, \quad \forall i = 1, \dots, M, \quad (15)$$

<sup>2</sup> The optimal solutions found by SDCut all had rank 7 except for one solution of rank 5.



Fig. 2: Image segmentation results on the Berkeley dataset. The red and blue marker indicates the annotated foreground and background super-pixels, respectively.

where  $M$  is the number of images and  $N = \sum_{i=1}^M N_i$  is the total number of pixels over all images. The matrix  $\mathbf{A} = \mathbf{A}_b + \frac{\mu}{N} \mathbf{A}_w$ , where  $\mathbf{A}_w$  is the intra-image affinity matrix and  $\mathbf{A}_b$  is the inter-image discriminative clustering cost matrix computed using the  $\chi^2$  distance between SIFT features in different images (see [26] for a details).

To solve this problem with BCR, we re-write (15) in the form (2) to obtain

$$\begin{aligned}
 & \underset{\mathbf{X} \in \mathbb{R}^{N \times r}}{\text{minimize}} && \text{tr}(\mathbf{X}^T \mathbf{A} \mathbf{X}) \\
 & \text{subject to:} && \text{tr}(\mathbf{X}^T \mathbf{Z}_i \mathbf{X}) = 1, \quad \forall i = 1, \dots, N \\
 & && \text{tr}(\mathbf{X}^T \mathbf{\Delta}_i \mathbf{X}) \leq \lambda^2, \quad \forall i = 1, \dots, M,
 \end{aligned} \tag{16}$$

where  $\mathbf{\Delta}_i = \delta_i \delta_i^T$  and  $\mathbf{Z}_i = \mathbf{e}_i \mathbf{e}_i^T$ . Finally, (16) is solved using BCR (4), following which one can recover the optimal score vector  $\mathbf{x}_p^*$  as the leading eigenvector of  $\mathbf{X}^*$ . The final binary solution is extracted by thresholding  $\mathbf{x}_p^*$  to obtain integer-valued labels [21].

**Experiments:** We compare BCR to two well-known co-segmentation methods, namely low-rank factorization [21] (denoted LR) and SDCut [19]. We use publicly available code for LR and SDCut. We test on the Weizman horses<sup>3</sup> and MSRC<sup>4</sup> datasets with a total of four classes (horse, car-front, car-back, and face) containing 6 ~ 10 images per class. Each image is over-segmented to 400 ~ 700 SLIC superpixels using the VLFeat [43] toolbox, giving a total of around 4000 ~ 7000 super-pixels per class. Relative to image segmentation problems, this application requires  $10 \times$  more variables.

Qualitative results are presented in Figure 3 while Table 2 provides a quantitative comparison. From Table 2, we observe that on average our method converges  $\sim 9.5 \times$  faster than SDCut and  $\sim 60 \times$  faster than LR. Moreover, the optimal objective value achieved by BCR is significantly lower than that achieved by both SDCut and LR methods. Figure 3 displays the visualization of the final score vector  $\mathbf{x}_p^*$  for selected images, depicting that in general SDCut and BCR produce similar results. Furthermore,

<sup>3</sup> [www.msri.org/people/members/eranb/](http://www.msri.org/people/members/eranb/)

<sup>4</sup> [www.research.microsoft.com/en-us/projects/objectclassrecognition/](http://www.research.microsoft.com/en-us/projects/objectclassrecognition/)

Table 1: Results on image segmentation. Numbers are the mean over the images in Fig. 2. Lower numbers are better. The proposed algorithm and the best performance are highlighted.

Method	BNCut	SDCut	BCR
Time (s)	<b>0.08</b>	27.64	0.97
Objective	10.84	6.40	<b>6.34</b>
Rank	1	7	2

Table 2: Co-segmentation results. The proposed algorithm and the best performance is highlighted.

Dataset		Test Cases			
		horse	face	car-back	car-front
Number of images		10	10	6	6
Variables in BQPs		4587	6684	4012	4017
Time (s)	LowRank	2647	1614	724	749
	SDCut	220	274	180	590
	<b>BCR</b>	<b>18.8</b>	<b>61.8</b>	<b>46.7</b>	<b>44.7</b>
Objective	LowRank	4.84	4.48	5.00	4.17
	SDCut	5.24	4.94	4.53	4.27
	<b>BCR</b>	<b>4.64</b>	<b>3.29</b>	<b>4.36</b>	<b>3.94</b>
Rank	LowRank	18	11	7	10
	SDCut	3	3	3	3
	<b>BCR</b>	<b>2</b>	<b>2</b>	<b>2</b>	<b>2</b>

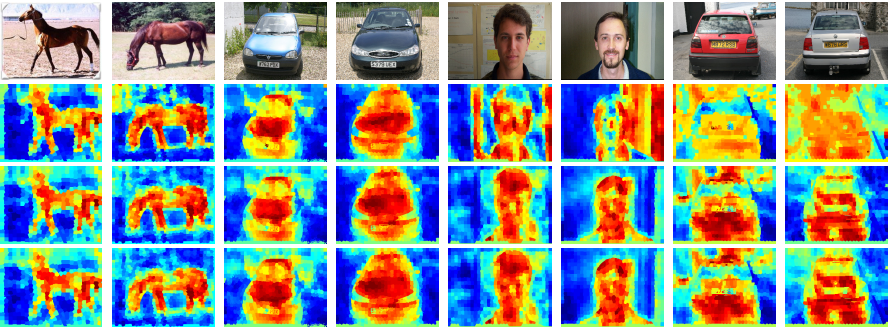


Fig. 3: Co-segmentation results on the Weizman horses and MSRC datasets. From top to bottom: the original images, the results of LR, SDCut, and BCR, respectively.

the optimal BCR score vector  $\mathbf{x}_p^*$  is extracted from a rank-2 solution, as compared to rank-3 and rank-7 solutions needed to get comparable results with SDCut and LR.

#### 4.4 Metric Learning on Manifolds

Large SDPs play a central role in manifold methods for classification and dimensionality reduction on image sets and videos [23,44,22]. Manifold methods rely heavily on covariance matrices, which accurately characterize second-order statistics of variation between images. Typical methods require computing distances between matrices along a Riemannian manifold—a task that is expensive for large matrices and limits the applicability of these techniques. It is of interest to perform dimensionality reduction on SPD matrices, thus enabling the use of covariance methods on very large problems.

In this section, we discuss dimensionality reduction on manifolds of SPD matrices using BCR, which is computationally much faster than the state-of-the-art while achieving comparable (and often better) performance. Consider a set of high-dimensional SPD matrices  $\{\mathbf{S}_1, \dots, \mathbf{S}_n\}$  where  $\mathbf{S}_i \in \mathcal{S}_{N \times N}^+$ . We can project these onto a low-dimensional

manifold of rank  $K < N$  by solving

$$\begin{aligned}
& \underset{\mathbf{X} \in S_{N \times N}^+, \eta_{ij} \geq 0}{\text{minimize}} && \text{tr}(\mathbf{X}) + \mu \sum_{i,j} \eta_{ij} \\
& \text{subject to} && \mathbb{D}_X(\mathbf{S}_i, \mathbf{S}_j) \leq u + \eta_{ij}, \quad \forall (i, j) \in \mathcal{C} \\
& && \mathbb{D}_X(\mathbf{S}_i, \mathbf{S}_j) \geq l - \eta_{ij}, \quad \forall (i, j) \in \mathcal{D}
\end{aligned} \tag{17}$$

where  $\mathbf{X}$  is a (low-dimensional) SPD matrix,  $\mathbb{D}_X$  is Riemannian distance metric, and  $\eta_{ij}$  are slack variables. The sets  $\mathcal{C}$  and  $\mathcal{D}$  contain pairs of similar/dissimilar matrices labeled by the user, and the scalars  $u$  and  $l$  are given upper and lower bounds. For simplicity, we measure distance using the log-Euclidean metric (LEM) defined by [22]

$$\mathbb{D}(\mathbf{S}_i, \mathbf{S}_j) = \|\log(\mathbf{S}_i) - \log(\mathbf{S}_j)\|_F^2 = \text{tr}((\mathbf{R}_i - \mathbf{R}_j)^T (\mathbf{R}_i - \mathbf{R}_j)), \tag{18}$$

where  $\mathbf{R}_i = \log(\mathbf{S}_i)$  is a matrix logarithm. When  $\mathbf{X}$  has rank  $K$ , it is a transformation onto the space of rank  $K$  covariance matrices, where the new distance is given by [22]

$$\mathbb{D}_X(\mathbf{S}_i, \mathbf{S}_j) = \text{tr}(\mathbf{X}(\mathbf{R}_i - \mathbf{R}_j)^T (\mathbf{R}_i - \mathbf{R}_j)). \tag{19}$$

We propose to solve the semi-definite program (17) using the representation  $\mathbf{X} = \mathbf{Y}\mathbf{Y}^T$  which puts our problem in the form (2) with  $\mathbf{A}_{ij} = (\mathbf{R}_i - \mathbf{R}_j)^T (\mathbf{R}_i - \mathbf{R}_j)$ . This problem is then solved using BCR, where the slack variables  $\{\eta_{ij}\}$  are removed and instead a hinge loss penalty approximately enforces the inequality constraints in (4). In our experiments we choose  $u = \rho - \xi\tau$  and  $l = \rho + \xi\tau$ , where  $\rho$  and  $\tau$  are the mean and standard deviation of the pairwise distances between  $\{S_i\}$  in the original space, respectively. The quantities  $\xi$  and  $\mu$  are treated as hyper-parameters.

**Experiments:** We analyze the performance of our approach (short BCRML) against state-of-the-art manifold metric learning algorithms using three image set classification databases: ETH-80, YouTube Celebrities (YTC), and YouTube Faces (YTF) [45]. The ETH-80 database consists of a 10 image set for each of 8 object categories. YTC contains 1,910 video sequences for 47 subjects from YouTube. YTF is a face verification database containing 3,425 videos of 1,595 different people. Features were extracted from images as described in [22]. Faces were cropped from each dataset using bounding boxes, and scaled to size  $20 \times 20$  for the ETH and YTC datasets. For YTF we used a larger  $30 \times 30$  scaling, as larger images were needed to replicate the results reported in [22].

We compare BCR to three state-of-the-art schemes: LEML [22] is based on a log-Euclidean metric, and minimizes the logdet divergence between matrices using Bregman projections. SPDML [23] optimizes a cost function on the Grassmannian manifold while making use of either the affine-invariant metric (AIM) or Stein metric. We use publicly available code for LEML and SPDML and follow the details in [23,22] to select algorithm specific hyper-parameters using cross-validation. For BCRML, we fix  $\alpha$  to be  $1/\sqrt{|\mathcal{C} \cup \mathcal{D}|}$  and  $\mu$  as  $\alpha/2$ . The  $\xi$  is fixed to 0.5, which performed well under cross-validation. For SPDML, the dimensionality of the target manifold  $K$  is fixed to 100. In LEML, the dimension cannot be reduced and thus the final dimension is the same as the original. Hence, for a fair comparison, we report the performance of BCRML using full target dimension (BCRML-full) as well as for  $K = 100$  (BCRML-100).

Table 3 summarizes the classification performance on the above datasets. We observe that BCRML performs almost the same or better than other ML algorithms. One can

Table 3: Image set classification results for state-of-the-art metric learning algorithms. The last three columns report computation time in seconds. The last 3 rows report performance using CDL-LDA after dimensionality reduction. Methods using the proposed BCR are listed in bold.

Method	ETH-80	YTC	YTF	Train (s)	Test (s)	Total (s)
AIM	89.25 $\pm$ 1.69	62.77 $\pm$ 2.89	59.82 $\pm$ 1.63	-	5.189	1463.3
Stein	89.00 $\pm$ 2.42	62.02 $\pm$ 2.71	57.56 $\pm$ 2.17	-	3.593	1013.3
LEM	90.00 $\pm$ 2.64	62.06 $\pm$ 3.04	59.78 $\pm$ 1.69	-	1.641	462
SPDML-AIM [23]	91.00 $\pm$ 3.39	65.32 $\pm$ 2.77	61.64 $\pm$ 1.46	3941	0.227	4005
SPDML-Stein [23]	90.75 $\pm$ 3.34	66.10 $\pm$ 2.92	61.66 $\pm$ 2.09	1447	<b>0.024</b>	1453.7
LEML [22]	92.00 $\pm$ 2.18	62.13 $\pm$ 3.13	60.92 $\pm$ 1.95	93	1.222	437.7
<b>BCRML-full</b>	92.00 $\pm$ 3.12	64.40 $\pm$ 2.92	60.58 $\pm$ 1.75	189	1.222	669.7
<b>BCRML-100</b>	92.25 $\pm$ 3.78	64.61 $\pm$ 2.65	<b>62.42 <math>\pm</math> 2.14</b>	<b>45</b>	0.291	127
CDL-LDA [44]	<b>94.25 <math>\pm</math> 3.36</b>	72.94 $\pm$ 1.81	N/A	-	1.073	302.7
LEML+CDL-LDA [22]	94.00 $\pm$ 3.57	73.01 $\pm$ 1.67	N/A	93	0.979	369
<b>BCRML-100+CDL-LDA</b>	93.75 $\pm$ 3.58	<b>73.48 <math>\pm</math> 1.83</b>	N/A	45	0.045	<b>57.7</b>

apply other algorithms to gain a further performance boost after projecting onto the low-dimensional manifold. Hence, we also provide a performance evaluation for LEML and BCRML using the LEM based CDL-LDA recognition algorithm [44]. The last three columns of Table 3 display the runtime measured on the YTC dataset. We note that BCRML-100 trains roughly  $2\times$  faster and overall runs about  $3.5\times$  faster than the next fastest method. Moreover, on testing using CDL-LDA, the overall computation time is approximately  $5\times$  faster in comparison to the next-best performing approach.

## 5 Conclusion

We have presented a novel biconvex relaxation framework (BCR) that enables the solution of general semidefinite programs (SDPs) at low complexity and with a small memory footprint. We have provided an alternating minimization (AM) procedure along with a new initialization method that, together, are guaranteed to converge, computationally efficient (even for large-scale problems), and able to handle a variety of SDPs. Comparisons of BCR with state-of-the-art methods for specific computer vision problems, such as segmentation, co-segmentation, and metric learning, show that BCR provides similar or better solution quality with significantly lower runtime. While this paper only shows applications for a select set of computer vision problems, determining the efficacy of BCR for other problems in signal processing, machine learning, control, etc. is left for future work.

**Acknowledgements:** The work of S. Shah and T. Goldstein was supported in part by the US National Science Foundation (NSF) under grant CCF-1535902 and by the US Office of Naval Research under grant N00014-15-1-2676. The work of A. Yadav and D. Jacobs was supported by the US NSF under grants IIS-1526234 and IIS-1302338. The work of C. Studer was supported in part by Xilinx Inc., and by the US NSF under grants ECCS-1408006 and CCF-1535897.

## References

1. Shi, J., Malik, J.: Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **22**(8) (2000) 888–905
2. Keuchel, J., Schno, C., Schellewald, C., Cremers, D.: Binary partitioning, perceptual grouping, and restoration with semidefinite programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(11) (2003) 1364–1379
3. Torr, P.H.: Solving markov random fields using semi definite programming. In: *Artificial Intelligence and Statistics*. (2003)
4. Goemans, M.X., Williamson, D.P.: Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)* **42**(6) (1995) 1115–1145
5. Arie-Nachimson, M., Kovalsky, S.Z., Kemelmacher-Shlizerman, I., Singer, A., Basri, R.: Global motion estimation from point matches. In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT)*, 2012 Second International Conference on, IEEE (2012) 81–88
6. Weinberger, K.Q., Saul, L.K.: Unsupervised learning of image manifolds by semidefinite programming. *International Journal of Computer Vision* **70**(1) (2006) 77–90
7. Mitra, K., Sheorey, S., Chellappa, R.: Large-scale matrix factorization with missing data under additional constraints. In: *Advances in Neural Information Processing Systems (NIPS)*. (2010) 1651–1659
8. Luo, Z.Q., Ma, W.k., So, A.M.C., Ye, Y., Zhang, S.: Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine* **27**(3) (May 2010) 20–34
9. Lasserre, J.B.: An explicit exact sdp relaxation for nonlinear 0-1 programs. In: *Integer Programming and Combinatorial Optimization*. Springer (2001) 293–303
10. Boyd, S., Vandenberghe, L.: Semidefinite programming relaxations of non-convex problems in control and combinatorial optimization. In: *Communications, Computation, Control, and Signal Processing*. Springer (1997) 279–287
11. Ecker, A., Jepson, A.D., Kutulakos, K.N.: Semidefinite programming heuristics for surface reconstruction ambiguities. In: *Computer Vision–ECCV 2008*. Springer (2008) 127–140
12. Shirdhonkar, S., Jacobs, D.W.: Non-negative lighting and specular object recognition. In: *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. Volume 2., IEEE* (2005) 1323–1330
13. Vandenberghe, L., Boyd, S.: Semidefinite programming. *SIAM review* **38**(1) (July 1996) 49–95
14. Heiler, M., Keuchel, J., Schnörr, C.: Semidefinite clustering for image segmentation with a-priori knowledge. In: *Pattern Recognition*. Springer (2005) 309–317
15. Shen, C., Kim, J., Wang, L.: A scalable dual approach to semidefinite metric learning. In: *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on, IEEE (2011) 2601–2608
16. Netrapalli, P., Jain, P., Sanghavi, S.: Phase retrieval using alternating minimization. In: *Advances in Neural Information Processing Systems (NIPS)*. (2013) 2796–2804
17. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via wirtinger flow: Theory and algorithms. *Information Theory, IEEE Transactions on* **61**(4) (2015) 1985–2007
18. Lee, J.D., Recht, B., Srebro, N., Tropp, J., Salakhutdinov, R.R.: Practical large-scale optimization for max-norm regularization. In: *Advances in Neural Information Processing Systems*. (2010) 1297–1305
19. Wang, P., Shen, C., van den Hengel, A.: A fast semidefinite approach to solving binary quadratic problems. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (June 2013)

20. Maji, S., Vishnoi, N.K., Malik, J.: Biased normalized cuts. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE (2011) 2057–2064
21. Journée, M., Bach, F., Absil, P.A., Sepulchre, R.: Low-rank optimization on the cone of positive semidefinite matrices. *SIAM Journal on Optimization* **20**(5) (2010) 2327–2351
22. Huang, Z., Wang, R., Shan, S., Li, X., Chen, X.: Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In: Proceedings of the 32nd International Conference on Machine Learning (ICML-15). (2015) 720–729
23. Harandi, M.T., Salzmann, M., Hartley, R.: From manifold to manifold: Geometry-aware dimensionality reduction for spd matrices. In: European Conference on Computer Vision, Springer (2014) 17–32
24. Bai, X., Yu, H., Hancock, E.R.: Graph matching using spectral embedding and alignment. In: International Conference on Pattern Recognition (ICPR). Volume 3., IEEE (2004) 398–401
25. Wang, C., Komodakis, N., Paragios, N.: Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding* **117**(11) (2013) 1610–1627
26. Joulin, A., Bach, F., Ponce, J.: Discriminative clustering for image co-segmentation. In: Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR). (2010)
27. Wang, P., Shen, C., van den Hengel, A.: Efficient SDP inference for fully-connected CRFs based on low-rank decomposition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (June 2015)
28. Schellewald, C., Schnörr, C.: Probabilistic subgraph matching based on convex relaxation. In: Energy minimization methods in computer vision and pattern recognition, Springer (2005) 171–186
29. Olsson, C., Eriksson, A.P., Kahl, F.: Solving large scale binary quadratic problems: Spectral methods vs. semidefinite programming. In: Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, IEEE (2007) 1–8
30. Lang, K.: Fixing two weaknesses of the spectral method. In: Advances in Neural Information Processing Systems (NIPS). (2005) 715–722
31. Sturm, J.F.: Using sedumi 1.02, a matlab toolbox for optimization over symmetric cones. *Optimization methods and software* **11**(1-4) (1999) 625–653
32. Toh, K.C., Todd, M., Tutuncu, R.: Sdpt3 - a matlab software package for semidefinite programming. *Optimization Methods and Software* **11** (1998) 545–581
33. Okatani, T., Deguchi, K.: On the wiberg algorithm for matrix factorization in the presence of missing components. *International Journal of Computer Vision* **72**(3) (2007) 329–337
34. Burer, S., Monteiro, R.D.: A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming* **95**(2) (2003) 329–357
35. Duchi, J.C., Singer, Y.: Efficient online and batch learning using forward backward splitting. *Journal of Machine Learning Research* **10** (2009) 2899–2934
36. Douglas, J., Gunn, J.E.: A general formulation of alternating direction methods. *Numerische Mathematik* **6**(1) (1964) 428–453
37. Xu, Y., Yin, W.: A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on imaging sciences* **6**(3) (2013) 1758–1789
38. Zheng, Q., Lafferty, J.: A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. In: Neural Information Processing Systems (NIPS). (2015)
39. Goldstein, T., Studer, C., Baraniuk, R.: A field guide to forward-backward splitting with a FASTA implementation. *arXiv eprint* **abs/1411.3406** (2014)
40. Wang, S., Siskind, J.M.: Image segmentation with ratio cut. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25** (2003) 675–690



41. Yu, S.X., Shi, J.: Segmentation given partial grouping constraints. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **26**(2) (2004) 173–183
42. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: *Proc. 8th Int'l Conf. Computer Vision. Volume 2. (July 2001)* 416–423
43. Vedaldi, A., Fulkerson, B.: Vlfeat: An open and portable library of computer vision algorithms (2008) (2012)
44. Wang, R., Guo, H., Davis, L.S., Dai, Q.: Covariance discriminative learning: A natural and efficient approach to image set classification. In: *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE (2012)* 2496–2503
45. Wolf, L., Hassner, T., Maoz, I.: Face recognition in unconstrained videos with matched background similarity. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, IEEE (2011)* 529–534