Biconvex Relaxation for Semidefinite Programming in Computer Vision

Sohil Shah*[,1], Abhay K. Yadav*[,1], Carlos D. Castillo[1], David W. Jacobs[1], Christoph Studer[2] and Tom Goldstein[1]
[1]University of Maryland, College Park; [2]Cornell University

ECCV'16 EUROPEAN CONFERENCE ON COMPUTER VISION

Motivation

- Many important problems requires solving for low rank SDPs with PSD constraint matrices.
- SDPs are commonly used in binary valued labeling problems (BQP) using semidefinite relaxation (SDR).
- In BQP, SDR transforms non-convex rank-1 constraint to PSD matrices with an all-ones diagonals.
- Existing solvers are **problem specific** or **slow.**
- Address challenges of memory and time; more general solver.

Contribution

- A novel framework to approximately solve SDPs with PSD constraint matrices efficiently.
- Effective initialization.
- 4x to 35x speedups with state-of-the-art results.

Overview

General SDPs

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

Specialized Solvers
- Interior Point Method
- Convex Algorithm
  - Spectral sub gradient method
  - Dual gradient descent
- Non convex Algorithm
  - Low rank approximation
  - Augmented Lagrangian

$\mathbf{Y} = \mathbf{X}\mathbf{X}^T$

Non-convex formulation

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

$$\underset{\mathbf{X}\in\mathbb{R}^{N\times r}}{\text{minimize}} \quad \text{tr}(\mathbf{X}^T\mathbf{C}\mathbf{X})$$
$$\text{subject to} \quad \mathbf{Q}_i = \mathbf{L}_i\mathbf{X}, \quad \|\mathbf{Q}_i\|_F^2 = b_i,$$
$$\mathbf{Q}_j = \mathbf{L}_j\mathbf{X}, \quad \|\mathbf{Q}_j\|_F^2 \le b_j,$$

Where $\mathbf{A} = \mathbf{L}^T\mathbf{L}$ and $\text{tr}(\mathbf{X}^T\mathbf{A}\mathbf{X}) = \text{tr}(\mathbf{X}^T\mathbf{L}^T\mathbf{L}\mathbf{X}) = \|\mathbf{L}\mathbf{X}\|_F^2$

Biconvex Relaxation (BCR) Framework

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

concave penalty counters relaxed equality constraints

where $\alpha > \beta > 0$ are relaxation parameters.

- Generally, $\alpha = 2\beta$ and $\beta = \|\mathbf{C}\|_2$
- β is chosen to match the curvature of objective with that of penalty term
- The problem is biconvex when $\mathbf{C} \in \mathcal{S}^{+}_{N\times N}$

Initialization

- If poorly initialized, Alternating Minimization algorithm may get trapped in local minima.
- Extend initialization algorithm previously used for phase retrieval problem [2,4].
- Decompose $\mathbf{C}$ into $\mathbf{C} = \mathbf{U}^T\mathbf{U}$ and use $\tilde{\mathbf{X}} = \mathbf{U}\mathbf{X}$.
- Then initializer $\tilde{\mathbf{X}}_0 = \lambda\mathbf{V}$, where λ and $\mathbf{V}$ are the r leading eigenvalue and eigenvectors of $\mathbf{Z} = \frac{1}{|\mathcal{E}|}\sum_{i\in\mathcal{E}} b_i\tilde{\mathbf{L}}_i^T\tilde{\mathbf{L}}_i.$
- Initialize $\mathbf{X}_0 = \mathbf{U}^{-1}\tilde{\mathbf{X}}_0$.

Optimization

- Alternating minimization Algorithm
  - **Stage 1**: Minimize w.r.t. $\{\mathbf{Q}_i\}$, i.e., minimize quadratic objective.
  - Project back into unit Frobenius-norm ball of radius $\sqrt{b_i}$
  - Expansion-reprojection update, $\mathbf{Q}_i \leftarrow \frac{\mathbf{L}_i\mathbf{X}}{\|\mathbf{L}_i\mathbf{X}\|_F}\min\left\{\sqrt{b_i}, \frac{\alpha}{\alpha-\beta_i}\|\mathbf{L}_i\mathbf{X}\|_F\right\}$
  - **Stage 2**: Minimize w.r.t. $\mathbf{X}$. Solves the least-squares problem,

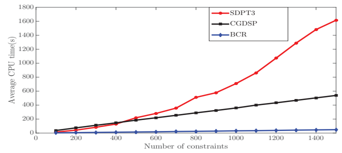
$$\mathbf{X} \leftarrow \left(\mathbf{C} + \alpha\sum_{i\in\{\mathcal{E}\cup\mathcal{B}\}}\mathbf{L}_i^T\mathbf{L}_i\right)^{-1}\left(\sum_{i\in\{\mathcal{E}\cup\mathcal{B}\}}\mathbf{L}_i^T\mathbf{Q}_i\right)$$

one time computation

Numerical Experiments

General Form Problems

Results on synthetic data for varying number of linear constraints



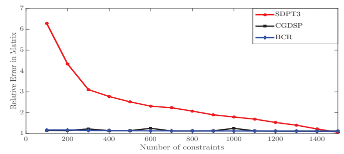
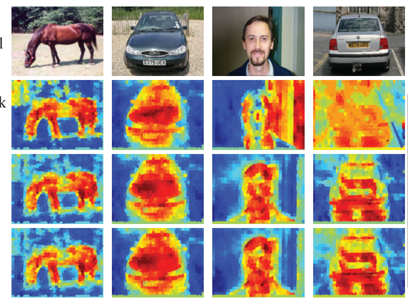


Image Segmentation

Berkeley dataset. The red and blue marker indicates annotated foreground and background super pixels

Method	BNCut	SDCut	**BCR**
Time (s)	**0.08**	27.64	0.97
Objective	10.84	6.40	**6.34**
Rank	1	7	2

Image Co-segmentation



Weizmann horses and MSRC datasets

		Test Cases			
Dataset		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
	BCR	**18.8**	**61.8**	**46.7**	**44.7**
Objective	LowRank	4.84	4.48	5.00	4.17
	SDCut	5.24	4.94	4.53	4.27
	BCR	**4.64**	**3.29**	**4.36**	**3.94**
Rank	LowRank	18	11	7	10
	SDCut	3	3	3	3
	BCR	**2**	**2**	**2**	**2**

Metric Learning on Manifolds

Image set classification results for state-of-the-art metric learning algorithms. Methods using the proposed BCR are listed in bold

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

References

[1] Wang, P., Shen, C., van den Hengel, A., A fast semidefinite approach to solving binary quadratic problems. CVPR 2013.
[2] Candes, E.J., Li, X., Soltanolkotabi, M., Phase retrieval via wirtinger flow: Theory and algorithms. IEEE Transaction on Information Theory, 2015.
[3] Zheng, Q., Lafferty, J., A convergent gradient descent algorithm for rank minimization and semidefinite programming from random linear measurements. NIPS 2015.
[4] Netrapalli, P., Jain, P., Sanghavi, S., Phase retrieval using alternating minimization. NIPS 2013.