



ICML 2014, Beijing

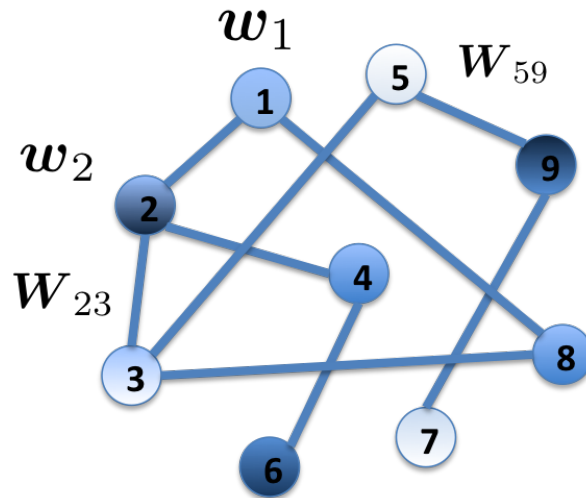
Scalable Semidefinite Relaxation for Maximum *A Posteriori* Estimation

Qixing Huang, Yuxin Chen, and Leonidas Guibas

Stanford University

Maximum A Posteriori (MAP) Inference

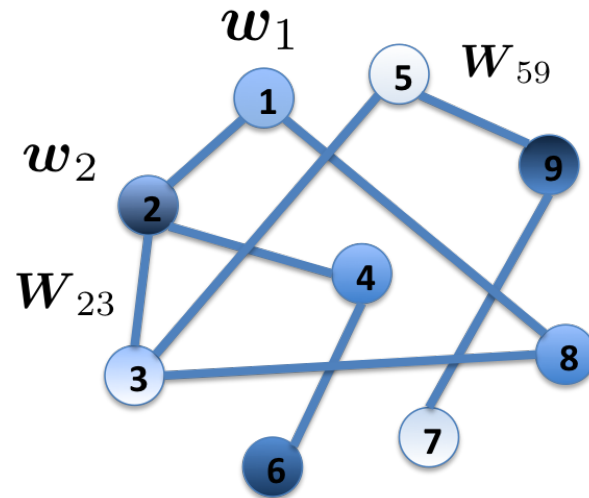
- Markov Random Field (MRF)



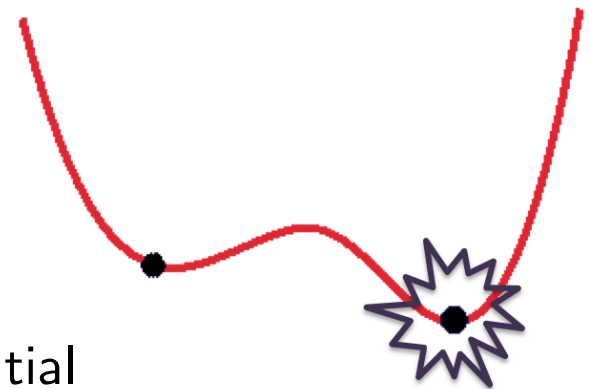
- w_i : potential function for vertices
- W_{ij} : potential function for edges

Maximum A Posteriori (MAP) Inference

- Markov Random Field (MRF)



- w_i : potential function for vertices
- W_{ij} : potential function for edges
- Maximum A Posteriori (MAP) Inference
 - Find **the mode** with the lowest energy / potential



A Large Number of Applications ...

- **Computer Vision Applications**




















- *Image Segmentation*
- *Geometric Surface Labeling*
- *Photo Montage*
- *Scene Decomposition*
- *Object Detection*
- *Color Segmentation*
- ...

- **Protein Folding**

- **Metric Labeling**

- **Error-Correcting Codes**

- ...

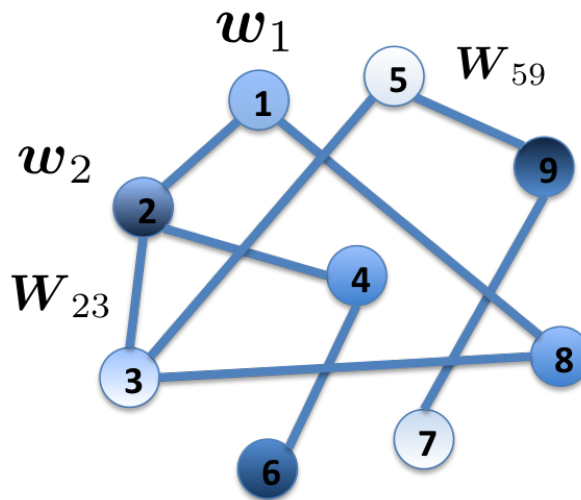
| | Variables | Labels | Order | Structure | Functions | Instances | Reference | Comment |
|---|-----------------|---------------|--------|------------|------------|-----------|-----------|--|
|  In-Painting (N4) J. Laitman et al. commented by J. Laitman and J. H. Rieger | 1400 | 4 | 2 | gM | gM | 2 | [40] | |
|  In-Painting (N2) J. Laitman et al. commented by J. Laitman and J. H. Rieger | 1400 | 4 | 2 | gM | gM | 2 | [40] | |
|  Color Segmentation (N4) J. Laitman et al. commented by J. Laitman and J. H. Rieger | 7800 | 3/2 | 2 | gM | gM | 2 | [40] | |
|  Color Segmentation (N2) J. Laitman et al. commented by J. Laitman and J. H. Rieger | 7800 | 3/2 | 2 | gM | gM | 2 | [40] | |
|  Color Segmentation K. Aizawa et al. commented by J. H. Rieger | 2100-4310 | 3/4 | 3 | gM | gM | 3 | [36] | |
|  Object Segmentation K. Aizawa et al. commented by J. H. Rieger | 8100 | 4/3 | 3 | gM | gM | 3 | [36] | |
|  MRF Photomontage M. Szeliski et al. commented by J. H. Rieger | 4300- 8100 | 3/1 | 3 | gM | gM | 3 | [34] | Visual noise not constant. |
|  MRF Stereo M. Szeliski et al. commented by J. H. Rieger | 11000 | 16/3 | 3 | gM | gM, gM, gM | 3 | [34] | |
|  MRF Inpainting M. Szeliski et al. commented by J. H. Rieger | 2100-8300 | 2/3 | 3 | gM | gM, gM, gM | 3 | [34] | |
|  Chinese Characters S. Nowozin et al. commented by S. Nowozin and J. H. Rieger | 490-1190 | 3 | 3 | gM | gM | 100 | [48] | |
|  Brain J. H. Rieger et al. commented by J. H. Rieger | 7800, 710-91 | 3 | 3 | gM | gM | 3 | [11] | Good example for large scale |
|  Scene Decomposition Gould et al. commented by S. Nowozin and J. H. Rieger | 18-28 | 3 | 3 | gM | gM | 191 | [21] | Use AUC@0.5. |
|  Geometric Surface Labeling (3) Galagher et al. commented by D. Zoran and J. H. Rieger | 24-113 | 3 | 3 | gM | gM | 300 | [23,25] | Use AUC@0.5. |
|  Geometric Surface Labeling (7) Galagher et al. commented by D. Zoran and J. H. Rieger | 24-113 | 7 | 3 | gM | gM | 300 | [23,25] | Use AUC@0.5. |
|  Matching N. Komodakis et al. commented by N. Komodakis and J. H. Rieger | 14-21 | 14-21 | 3 | Full gM | gM | 4 | [44] | Look for 2D/3D relation in the input. |
|  Cell Tracking E. X. Kautler et al. commented by E. X. Kautler | 4134 | 2 | 3 | gM | gM | 1 | [26] | Visual noise not constant. |
|  Image Segmentation E. Arino et al. commented by E. Arino | 100-214 | 100-214 | 2 | gM | gM | 100 | [10] | Trained to use per-class. Use AUC@0.5. |
|  Hierarchical image segmentation K. Sugrue et al. commented by K. Sugrue and J. H. Rieger | 100-601 | 100-601 | 34-601 | gM | gM | 710 | [28] | Trained to use per-class. Use AUC@0.5. |
|  3D Neuron Segmentation E. Arino et al. commented by E. Arino | 700-10100 | 700- 10100 | 2 | gM | gM | 2 | [12,13] | Trained to use per-class. Use AUC@0.5. |

OpenGM Benchmark

Problem Setup

- **Model**

- n vertices (x_1, \dots, x_n)
- m different states $\iff x_i \in \{1, \dots, m\}$



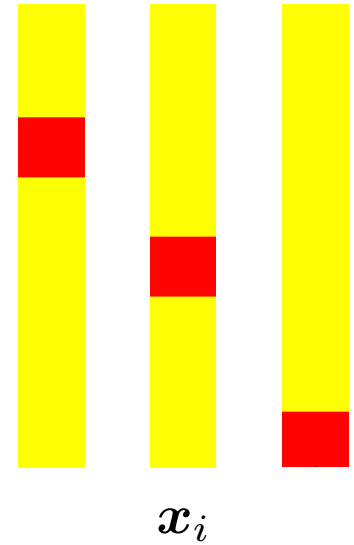
- **Goal:**

$$\begin{aligned} \text{maximize} \quad & \underbrace{f(x_1, \dots, x_n)}_{\text{negative energy function}} := \sum_{i=1}^n w_i(x_i) + \sum_{(i,j) \in \mathcal{G}} W_{ij}(x_i, x_j) \\ \text{s.t.} \quad & x_i \in \{1, \dots, m\} \end{aligned}$$

Matrix Representation

- Representation of Each x_i

- m possible states $\iff x_i \in \{e_1, e_2, \dots, e_m\}$



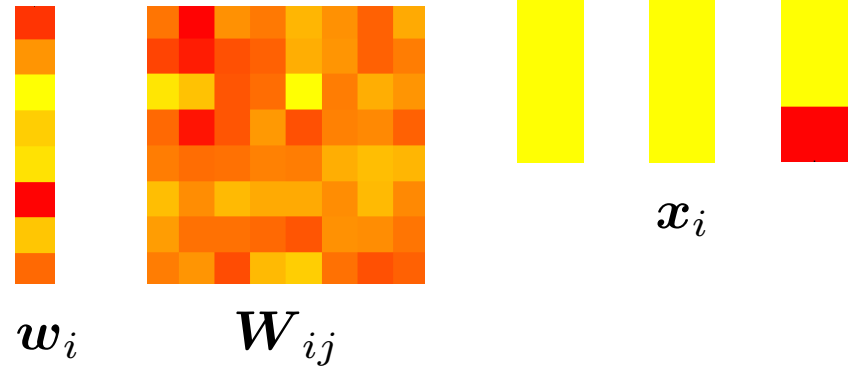
Matrix Representation

- Representation of Each x_i

- m possible states $\iff x_i \in \{e_1, e_2, \dots, e_m\}$

- Representation of Potentials

- potential on vertices: $w_i \in \mathbb{R}^m$
- potential on edges: $W_{ij} \in \mathbb{R}^{m \times m}$



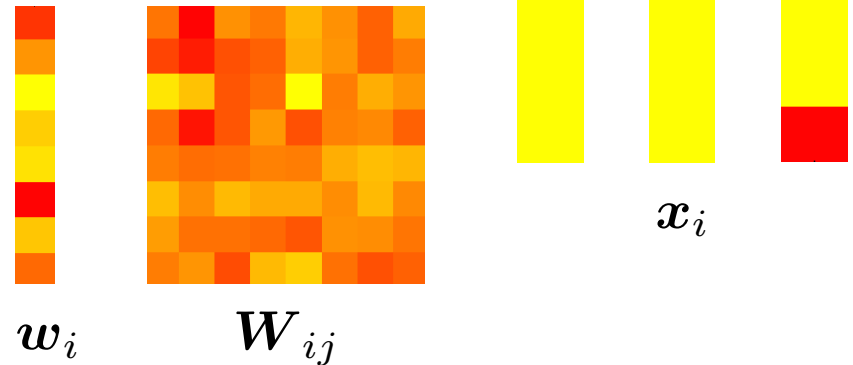
Matrix Representation

- Representation of Each x_i

- m possible states $\iff x_i \in \{e_1, e_2, \dots, e_m\}$

- Representation of Potentials

- potential on vertices: $w_i \in \mathbb{R}^m$
- potential on edges: $W_{ij} \in \mathbb{R}^{m \times m}$



- Equivalent Integer Program:

$$\begin{aligned} &\text{maximize} && f(x_1, \dots, x_n) := \sum_{i=1}^n \langle w_i, x_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle W_{ij}, x_i x_j^\top \rangle \\ &\text{s.t.} && x_i \in \{e_1, \dots, e_m\} \end{aligned}$$

- Non-Convex!

Matrix Representation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{x}_i \mathbf{x}_j^\top \rangle \\ \text{s.t.} \quad & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Matrix Representation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{x}_i \mathbf{x}_j^\top \rangle \\ \text{s.t.} \quad & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

• **Auxiliary Variable** $\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1n} \\ \mathbf{X}_{12}^\top & \mathbf{X}_{22} & \cdots & \mathbf{X}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{X}_{1n}^\top & \mathbf{X}_{2n}^\top & \cdots & \mathbf{X}_{nn} \end{bmatrix}$

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \mathbf{X}_{ij} = \mathbf{x}_i \mathbf{x}_j^\top, \quad \mathbf{X}_{ii} = \mathbf{x}_i \mathbf{x}_i^\top = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Matrix Representation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{x}_i \mathbf{x}_j^\top \rangle \\ \text{s.t.} \quad & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

• **Auxiliary Variables** $\mathbf{X} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} & \cdots & \mathbf{X}_{1n} \\ \mathbf{X}_{12}^\top & \mathbf{X}_{22} & \cdots & \mathbf{X}_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{X}_{1n}^\top & \mathbf{X}_{2n}^\top & \cdots & \mathbf{X}_{nn} \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{x} \mathbf{x}^\top, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Convex Relaxation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{x}\mathbf{x}^\top, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Convex Relaxation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \mathbf{X} = \mathbf{x}\mathbf{x}^\top, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

- Semidefinite Relaxation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0} \\ & \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Convex Relaxation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

Convex Relaxation

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\} \end{aligned}$$

- Relax the Constraints $\mathbf{x}_i \in \{\mathbf{e}_1, \dots, \mathbf{e}_m\}$

$$\begin{aligned} \text{maximize} \quad & f(\mathbf{x}_1, \dots, \mathbf{x}_n) := \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1, \\ & \mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i,j) \in \mathcal{G} \end{aligned}$$

Our Semidefinite Formulation

- Final Semidefinite Program (SDR):

$$\begin{aligned}
 &\text{maximize} && \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\
 &\text{s.t.} && \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \quad \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\
 &&& \mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1, \\
 &&& \mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i,j) \in \mathcal{G}
 \end{aligned}$$

- Low-Rank and Sparse!*

- $O(nm^2)$ linear equality constraints

$$\mathbf{X} = \mathbf{x} \cdot \mathbf{x}^\top$$

Superiority to Linear Programming Relaxation

Semidefinite Relaxation (SDR)

$$\begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0},$$
$$\mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i)$$
$$\mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1$$
$$\mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G}$$

LP Relaxation

$$\mathbf{X}_{ij} \mathbf{1} = \mathbf{x}_i \quad (1 \leq i, j \leq n)$$
$$\mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i)$$
$$\mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1$$
$$\mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G}$$

- Shall we enforce the marginalization constraints? $(\underbrace{\mathbf{X}_{ij} \mathbf{1} = \mathbf{x}_i, \quad 1 \leq i, j \leq n}_{\Theta(n^2m) \text{ constraints}})$

Superiority to Linear Programming Relaxation

Semidefinite Relaxation (SDR)

$$\begin{bmatrix} \mathbf{1} & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0},$$
$$\mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i)$$
$$\mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1$$
$$\mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G}$$

LP Relaxation

$$\mathbf{X}_{ij} \mathbf{1} = \mathbf{x}_i \quad (1 \leq i, j \leq n)$$
$$\mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i)$$
$$\mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1$$
$$\mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G}$$

- Shall we enforce the marginalization constraints? $(\underbrace{\mathbf{X}_{ij} \mathbf{1} = \mathbf{x}_i, \quad 1 \leq i, j \leq n}_{\Theta(n^2 m)} \text{ constraints})$
- Answer: No!

Proposition

Any feasible solution to SDR necessarily satisfies $\mathbf{X}_{ij} \mathbf{1} = \mathbf{x}_i$.

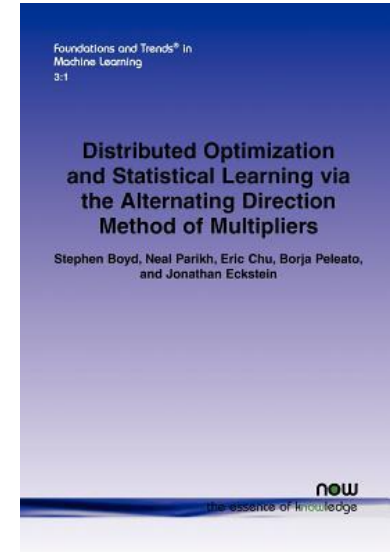
$O(nm^2)$ v.s. $O(n^2 m + nm^2)$ linear equality constraints!

ADMM

- Alternating Direction Methods of Multipliers
 - Fast convergence in the first several tens of iterations

Semidefinite Relaxation (SDR)

$$\begin{aligned} \max \quad & \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \\ & \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1, \\ & \mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G} \end{aligned}$$

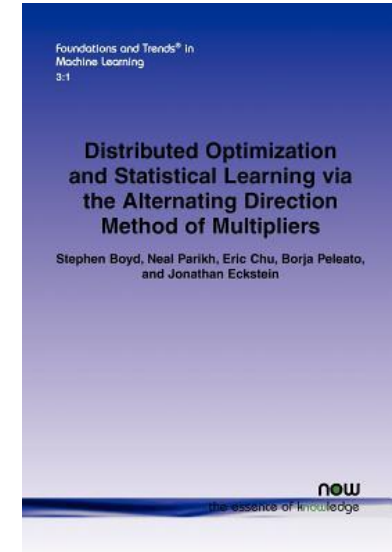


ADMM

- Alternating Direction Methods of Multipliers
 - Fast convergence in the first several tens of iterations

Semidefinite Relaxation (SDR)

$$\begin{aligned} \max \quad & \sum_{i=1}^n \langle \mathbf{w}_i, \mathbf{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \mathbf{W}_{ij}, \mathbf{X}_{ij} \rangle \\ \text{s.t.} \quad & \begin{bmatrix} 1 & \mathbf{x}^\top \\ \mathbf{x} & \mathbf{X} \end{bmatrix} \succeq \mathbf{0}, \\ & \mathbf{X}_{ii} = \text{diag}(\mathbf{x}_i) \\ & \mathbf{x}_i \geq \mathbf{0}, \quad \mathbf{1}^\top \mathbf{x}_i = 1, \\ & \mathbf{X}_{ij} \geq \mathbf{0}, \quad \forall (i, j) \in \mathcal{G} \end{aligned}$$



Generic Formulation

$$\begin{aligned} \max \quad & \langle \mathbf{C}, \mathbf{X} \rangle \\ \text{s.t.} \quad & \mathbf{A}(\mathbf{X}) = \mathbf{b}, \\ & \mathbf{B}(\mathbf{X}) \geq \mathbf{0}, \\ & \mathbf{X} \succeq \mathbf{0}. \end{aligned}$$

- $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are all highly sparse!

Scalability?

Generic Formulation

$$\begin{array}{ll} \max & \langle \mathbf{C}, \mathbf{X} \rangle & \text{dual vars} \\ \text{s.t.} & \mathcal{A}(\mathbf{X}) = \mathbf{b}, & \mathbf{y} \\ & \mathcal{B}(\mathbf{X}) \succeq \mathbf{0}, & \mathbf{z} \succeq \mathbf{0} \\ & \mathbf{X} \succeq \mathbf{0}. & \mathbf{S} \succeq \mathbf{0} \end{array}$$

- $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are all sparse!
 - All operations are fast except ...

Scalability?

Generic Formulation

$$\begin{array}{ll} \max & \langle \mathbf{C}, \mathbf{X} \rangle & \text{dual vars} \\ \text{s.t.} & \mathcal{A}(\mathbf{X}) = \mathbf{b}, & \mathbf{y} \\ & \mathcal{B}(\mathbf{X}) \succeq \mathbf{0}, & \mathbf{z} \succeq \mathbf{0} \\ & \mathbf{X} \succeq \mathbf{0}. & \mathbf{S} \succeq \mathbf{0} \end{array}$$

- $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are all sparse!
 - All operations are fast except ...

$$\mathbf{X}^{(t)} = \underbrace{\left(\mathbf{X}^{(t-1)} - \frac{\mathbf{C} + \mathcal{A}^*(\mathbf{y}^{(t)}) - \mathcal{B}^*(\mathbf{z}^{(t)})}{\mu} \right)}_{\text{projection onto PSD cone}} \succeq \mathbf{0}$$

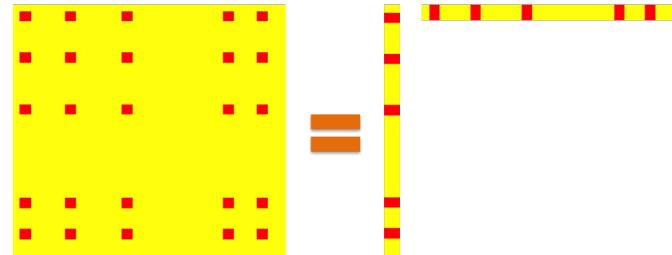
- Eigen-decomposition of dense matrices is expensive!

Accelerated ADMM (SDPAD-LR)

$$\mathbf{X}^{(t)} = \underbrace{\left(\mathbf{X}^{(t-1)} - \frac{\mathbf{C} + \mathcal{A}^*(\mathbf{y}^{(t)}) - \mathcal{B}^*(\mathbf{z}^{(t)})}{\mu} \right)}_{\text{projection onto PSD cone}} \succeq \mathbf{0}$$

- Recall: the ground truth obeys $\text{rank}(\mathbf{X}) = 1$

- Enforce / Exploit Low-Rank Structure!

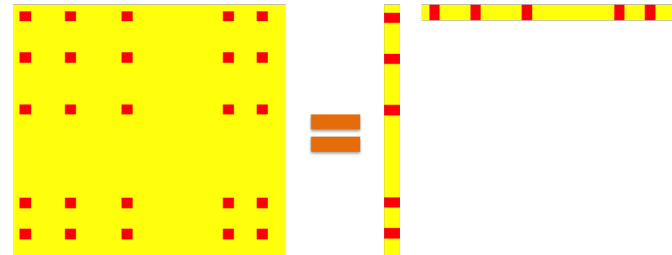


Accelerated ADMM (SDPAD-LR)

$$\mathbf{X}^{(t)} = \underbrace{\left(\mathbf{X}^{(t-1)} - \frac{\mathbf{C} + \mathcal{A}^*(\mathbf{y}^{(t)}) - \mathcal{B}^*(\mathbf{z}^{(t)})}{\mu} \right)}_{\text{projection onto PSD cone}} \gamma_0$$

- Recall: the ground truth obeys $\text{rank}(\mathbf{X}) = 1$

- Enforce / Exploit Low-Rank Structure!



- Our Strategy:

- Only keep *rank-r* approximation of $\mathbf{X}^{(t)} \approx \mathbf{Y}^{(t)} \mathbf{Y}^{(t)\top}$

eigens of $\left(\underbrace{\mathbf{Y}^{(t-1)} \mathbf{Y}^{(t-1)\top}}_{\text{low rank}} - \underbrace{\frac{\mathbf{C} + \mathcal{A}^*(\mathbf{y}^{(t)}) - \mathcal{B}^*(\mathbf{z}^{(t)})}{\mu}}_{\text{sparse}} \right)$

Accelerated ADMM (SDPAD-LR)

- **Our Strategy:**

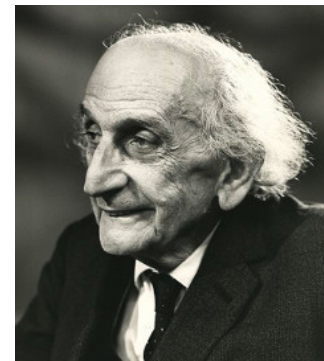
- Only keep *rank- r* approximation of $\mathbf{X}^{(t)} \approx \mathbf{Y}^{(t)}\mathbf{Y}^{(t)\top}$

eigens of $\left(\underbrace{\mathbf{Y}^{(t-1)}\mathbf{Y}^{(t-1)\top}}_{\text{low rank}} - \underbrace{\frac{\mathbf{C} + \mathcal{A}^*(\mathbf{y}^{(t)}) - \mathcal{B}^*(\mathbf{z}^{(t)})}{\mu}}_{\text{sparse}} \right)$

- Numerically fast

- e.g. **Lanczos Process** $O(nmr^2 + m^2|\mathcal{G}|)$

- Empirically, $r \approx 8$

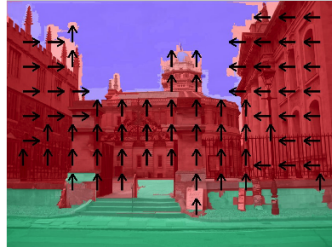


Cornelius Lanczos

Benchmark Data Sets

- Benchmark

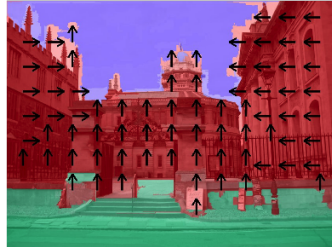
- OPENGM2
- PIC
- ORIENT



Benchmark Data Sets

- **Benchmark**

- OPENGM2
- PIC
- ORIENT

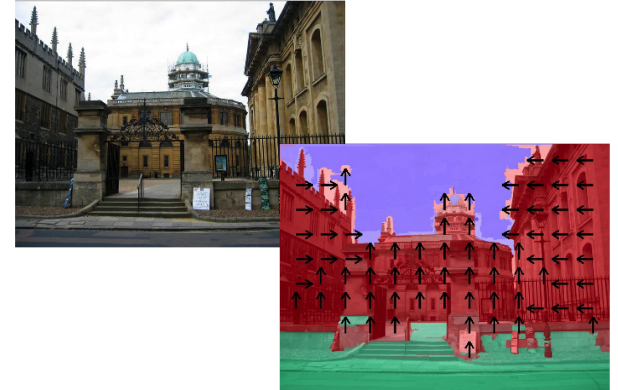


| categories | graphs | n | m | # instances | avg time |
|-------------|--------|--------|-------|-------------|----------|
| PIC-Object | full | 60 | 11-21 | 37 | 5m32s |
| PIC-Folding | mixed | 2K | 2-503 | 21 | 21m42s |
| PIC-Align | dense | 30-400 | 20-93 | 19 | 37m63s |
| GM-Label | sparse | 1K | 7 | 324 | 6m32s |
| GM-Char | sparse | 5K-18K | 2 | 100 | 1h13m |
| GM-Montage | grid | 100K | 5,7 | 3 | 9h32m |
| GM-Matching | dense | 19 | 19 | 4 | 2m21s |
| ORIENT | sparse | 1K | 16 | 10 | 10m21s |

All problems can be solved within reasonable time!

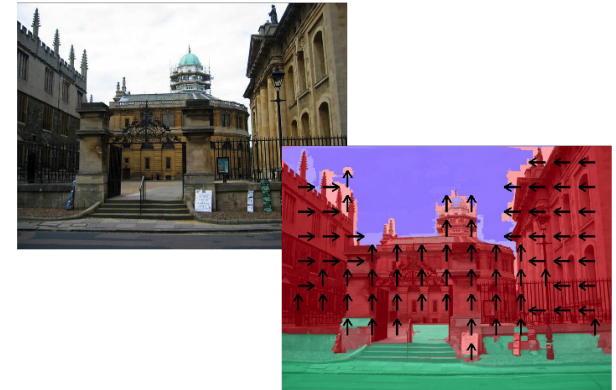
Empirical Convergence: Example

- Benchmark: Geometric Surface Labeling (gm275)
 - matrix size: 5201; # constraints: 218791
 - Stopping criterion: duality gap $< 10^{-3}$



Empirical Convergence: Example

- Benchmark: Geometric Surface Labeling (gm275)
 - matrix size: 5201; # constraints: 218791
 - Stopping criterion: duality gap $< 10^{-3}$



| | SDPAD-LR (our algorithm) | SDPAD (original ADMM by Wen'10) |
|---------------------------|-----------------------------|------------------------------------|
| time | 21:33 | 41:33:21 |
| duality gap | 5.1×10^{-4} | 1.2×10^{-4} |
| primal-dual infeasibility | 1.3×10^{-6} | 3.1×10^{-6} |

| | SDPNAL (ADMM w/ Newton-CG) | MOSEK interior point |
|---------------------------|-------------------------------|-------------------------|
| time | 21:34:35 | N/A |
| duality gap | 0.97×10^{-4} | N/A |
| primal-dual infeasibility | 4.5×10^{-7} | N/A |

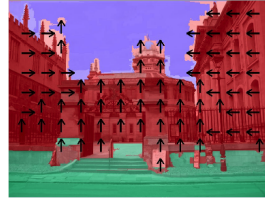
- *SDPAD-LR converges to the correct optimizer of SDP in these problems!*

Performance on MAP Problems

- Performance Measures

- mean objective values

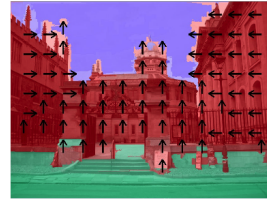
- the percentage of an algorithm achieving the best result



Performance on MAP Problems

- Performance Measures

- mean objective values

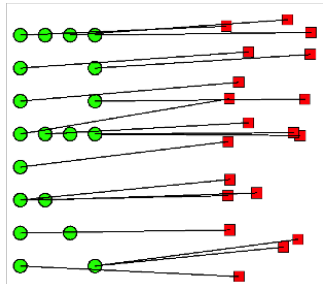


- the percentage of an algorithm achieving the best result

| | SDPAD-LR | Ficolofa | BRAOBB | α -expand | TRWS-LF2 | ogm-TRBP |
|-------------|----------------------------------|--------------------------------|-----------------------------|-------------------------------|--------------------|--------------------|
| ORIENT | -7834.6 100% | na | -3059.2 0% | -7695.4 0% | -7592.4 0% | -7553.8 0% |
| PIC-Object | -19316.12 97.3% | -19308.94 91.9% | -19113.87 24.3% | -10106.8 0% | -19020.82 59.5% | -18900.81 32.2% |
| PIC-Folding | -5963.68 100% | -5963.68 100% | -5927.01 42.9% | -5652.76 14.2% | -5905.01 38.1% | -5907.24 42.9% |
| PIC-Align | 2285.23 100% | 2285.34 90% | 2285.34 90% | 2285.34 90% | 2286.64 80% | 2289.12 70% |
| GM-Label | -476.95 100% | na | na | -476.95 100% | -476.95 99.67% | 486.42 40% |
| GM-Char | -59550.67 86.1% | na | na | na | -49519.44 11% | -49507.98 6% |
| GM-Montage | 168298.00 66.3% | na | na | 168220.00 33.3% | 735193.0 0% | 235611.00 0% |
| GM-Matching | 44.19 0% | na | 21.22 100% | na | 32.38 0% | 5.5e10 0% |

Implications from Empirical Results

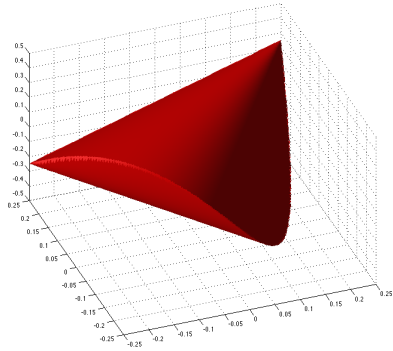
| | SDPAD-LR | Ficolofa | BRAOBB | α -expand | TRWS-LF2 | ogm-TRBP |
|-------------|----------------------------------|--------------------------------|-----------------------------|-------------------------------|--------------------|--------------------|
| ORIENT | -7834.6 100% | na | -3059.2 0% | -7695.4 0% | -7592.4 0% | -7553.8 0% |
| PIC-Object | -19316.12 97.3% | -19308.94 91.9% | -19113.87 24.3% | -10106.8 0% | -19020.82 59.5% | -18900.81 32.2% |
| PIC-Folding | -5963.68 100% | -5963.68 100% | -5927.01 42.9% | -5652.76 14.2% | -5905.01 38.1% | -5907.24 42.9% |
| PIC-Align | 2285.23 100% | 2285.34 90% | 2285.34 90% | 2285.34 90% | 2286.64 80% | 2289.12 70% |
| GM-Label | -476.95 100% | na | na | -476.95 100% | -476.95 99.67% | 486.42 40% |
| GM-Char | -59550.67 86.1% | na | na | na | -49519.44 11% | -49507.98 6% |
| GM-Montage | 168298.00 66.3% | na | na | 168220.00 33.3% | 735193.0 0% | 235611.00 0% |
| GM-Matching | 44.19 0% | na | 21.22 100% | na | 32.38 0% | 5.5e10 0% |



- SDPAD-LR is *top-performing* on all datasets
except *GM-Matching*

Concluding Remarks

- **Semidefinite Relaxation for MAP Inference**
 - SDP, if computationally feasible, outperforms many algorithms
 - Exploit underlying structure to accelerate SDP solvers
- **The Way Ahead**
 - Theoretical Support for Our Accelerated Algorithm (SDPAD-LR)



Thanks! Questions?

