ICML 2014, Beijing

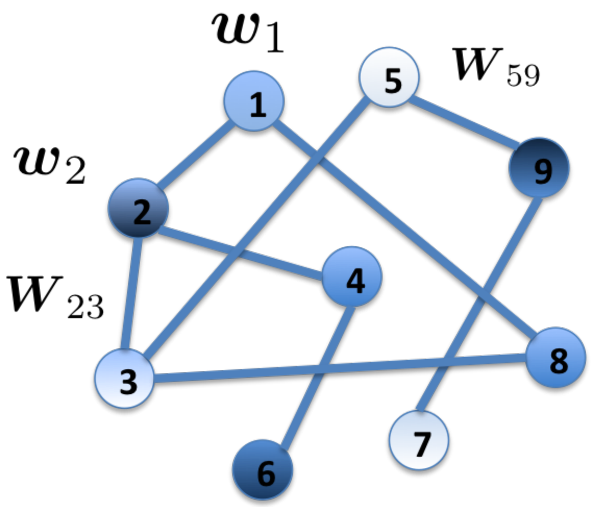
Scalable Semidefinite Relaxation for Maximum *A Posteriori* Estimation

Qixing Huang, Yuxin Chen, and Leonidas Guibas

Stanford University

Maximum A Posteriori (MAP) Inference

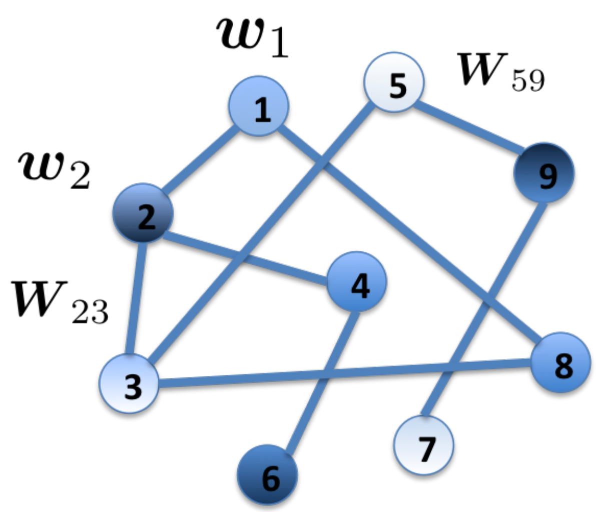
- Markov Random Field (MRF)



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

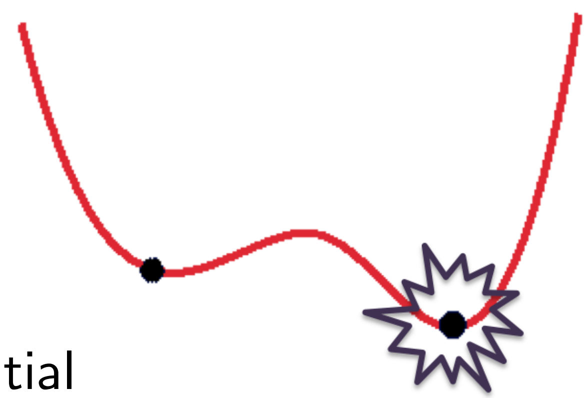
Maximum A Posteriori (MAP) Inference

- Markov Random Field (MRF)



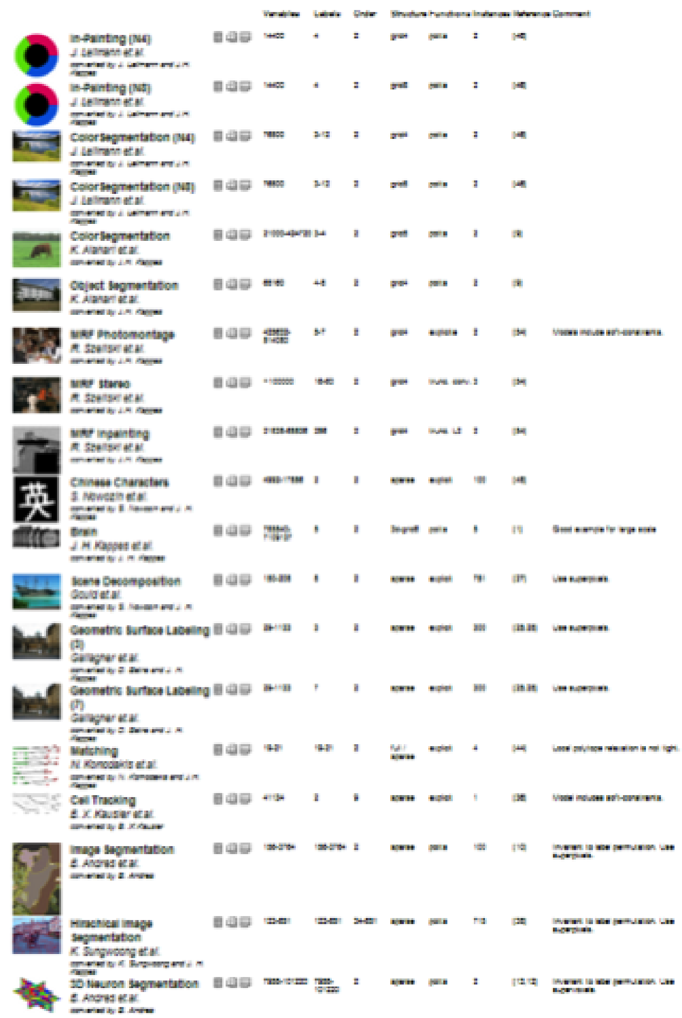
  - w_i: potential function for vertices
  - W_{ij}: potential function for edges

- Maximum A Posteriori (MAP) Inference
  - Find **the mode** with the lowest energy / potential



A Large Number of Applications ...

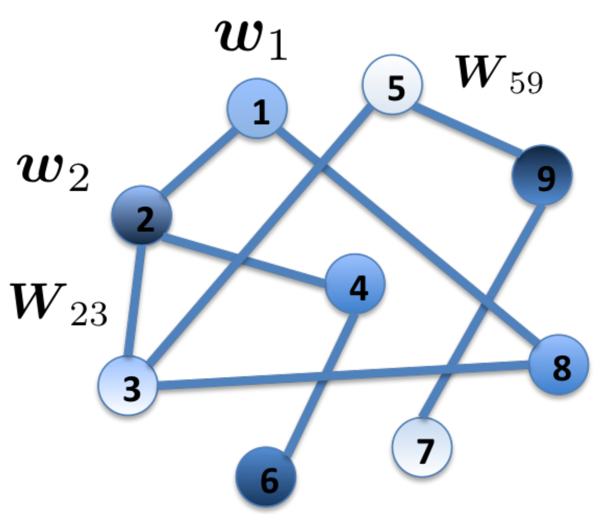
- **Computer Vision Applications**
  - *Image Segmentation*
  - *Geometric Surface Labeling*
  - *Photo Montage*
  - *Scene Decomposition*
  - *Object Detection*
  - *Color Segmentation*
  - ...
- **Protein Folding**
- **Metric Labeling**
- **Error-Correcting Codes**
- ...



OpenGM Benchmark

Problem Setup

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



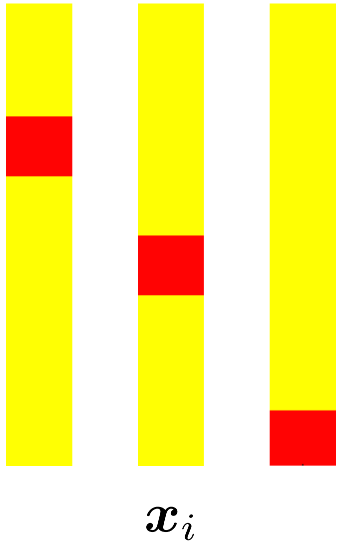
- **Goal:**

$$\text{maximize} \quad \underbrace{f(x_1, \cdots, x_n)}_{\text{negative energy function}} := \sum_{i=1}^{n} \boldsymbol{w}_i(x_i) + \sum_{(i,j)\in\mathcal{G}} \boldsymbol{W}_{ij}(x_i, x_j)$$

$$\text{s.t.} \quad x_i \in \{1, \cdots, m\}$$

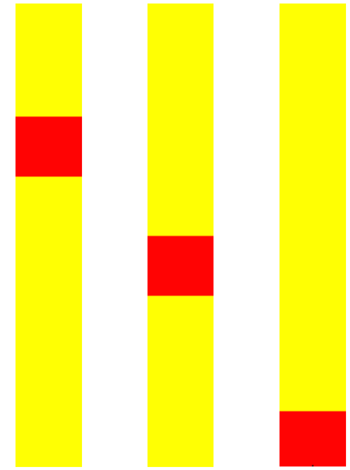
Matrix Representation

- **Representation of Each x_i**
  - m possible states $\iff$ $\boldsymbol{x}_i \in \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m\}$

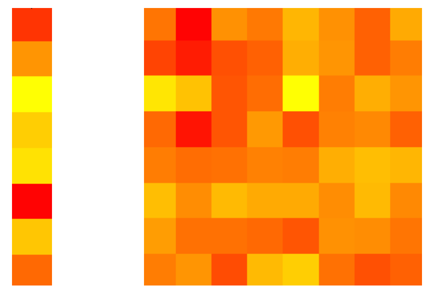


Matrix Representation

- **Representation of Each x_i**
  - m possible states $\quad\Longleftrightarrow\quad \boldsymbol{x}_i \in \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m\}$

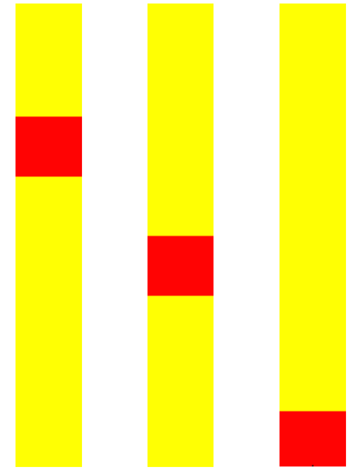


- **Representation of Potentials**
  - potential on vertices: $\boldsymbol{w}_i \in \mathbb{R}^m$
  - potential on edges: $\boldsymbol{W}_{ij} \in \mathbb{R}^{m \times m}$



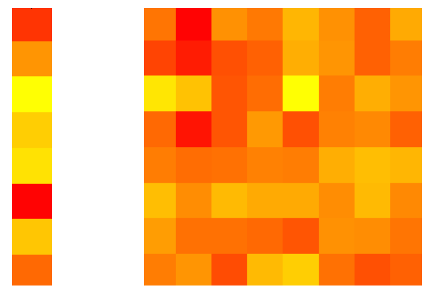
Matrix Representation

- **Representation of Each x_i**
  - m possible states $\Longleftrightarrow$ $\boldsymbol{x}_i \in \{\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_m\}$



$\boldsymbol{x}_i$

- **Representation of Potentials**
  - potential on vertices: $\boldsymbol{w}_i \in \mathbb{R}^m$
  - potential on edges: $\boldsymbol{W}_{ij} \in \mathbb{R}^{m \times m}$



$\boldsymbol{w}_i$ $\boldsymbol{W}_{ij}$

- **Equivalent Integer Program:**

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

  - Non-Convex!

Matrix Representation

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

Matrix Representation

maximize $\quad f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{x}_i \boldsymbol{x}_j^\top \rangle$

s.t. $\quad \boldsymbol{x}_i \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$

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

maximize $\quad f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle$

s.t. $\quad \boldsymbol{X}_{ij} = \boldsymbol{x}_i \boldsymbol{x}_j^\top, \quad \boldsymbol{X}_{ii} = \boldsymbol{x}_i \boldsymbol{x}_i^\top = \mathrm{diag}(\boldsymbol{x}_i)$

$\quad \boldsymbol{x}_i \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$

Matrix Representation

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

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

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

Convex Relaxation

maximize $\quad f\left(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n\right) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle$

s.t. $\quad \boldsymbol{X} = \boldsymbol{x}\boldsymbol{x}^\top, \quad \boldsymbol{X}_{ii} = \mathrm{diag}(\boldsymbol{x}_i)$

$\quad \boldsymbol{x}_i \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$

Convex Relaxation

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

- **Semidefinite Relaxation**

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

Convex Relaxation

maximize $\quad f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j) \in \mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle$

s.t. $\quad \begin{bmatrix} 1 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{bmatrix} \succeq \mathbf{0}, \quad \boldsymbol{X}_{ii} = \mathrm{diag}(\boldsymbol{x}_i)$

$\boldsymbol{x}_i \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$

Convex Relaxation

$$\text{maximize} \quad f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle$$

$$\text{s.t.} \quad \begin{bmatrix} 1 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{bmatrix} \succeq \boldsymbol{0}, \quad \boldsymbol{X}_{ii} = \mathrm{diag}(\boldsymbol{x}_i)$$

$$\boldsymbol{x}_i \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_m\}$$

- **Relax the Constraints** $x_i \in \{e_1, \cdots, e_m\}$

$$\text{maximize} \quad f(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) := \sum_{i=1}^{n} \langle \boldsymbol{w}_i, \boldsymbol{x}_i \rangle + \sum_{(i,j)\in\mathcal{G}} \langle \boldsymbol{W}_{ij}, \boldsymbol{X}_{ij} \rangle$$

$$\text{s.t.} \quad \begin{bmatrix} 1 & \boldsymbol{x}^\top \\ \boldsymbol{x} & \boldsymbol{X} \end{bmatrix} \succeq \boldsymbol{0}, \quad \boldsymbol{X}_{ii} = \mathrm{diag}(\boldsymbol{x}_i)$$

$$\boldsymbol{x}_i \geq \boldsymbol{0}, \quad \boldsymbol{1}^\top \boldsymbol{x}_{\mathbf{i}} = \boldsymbol{1},$$

$$\boldsymbol{X}_{ij} \geq \boldsymbol{0}, \quad \forall (i,j) \in \mathcal{G}$$

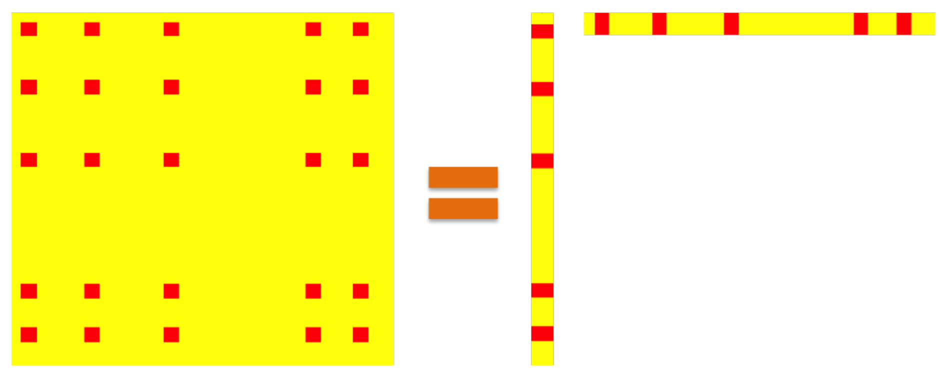
Our Semidefinite Formulation

- **Final Semidefinite Program (SDR):**

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

- *Low-Rank and Sparse!*

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



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

Superiority to Linear Programming Relaxation

Semidefinite Relaxation (SDR)

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

LP Relaxation

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

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

Superiority to Linear Programming Relaxation

Semidefinite Relaxation (SDR)

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

LP Relaxation

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

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

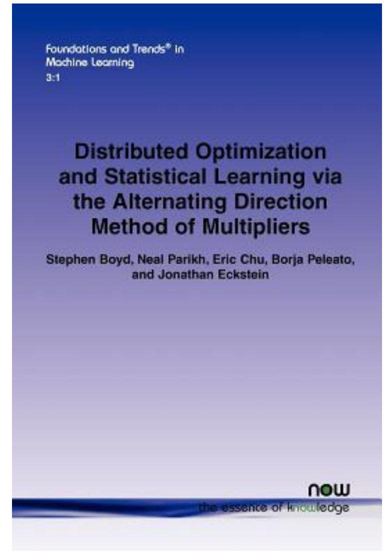
Proposition

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

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

ADMM

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

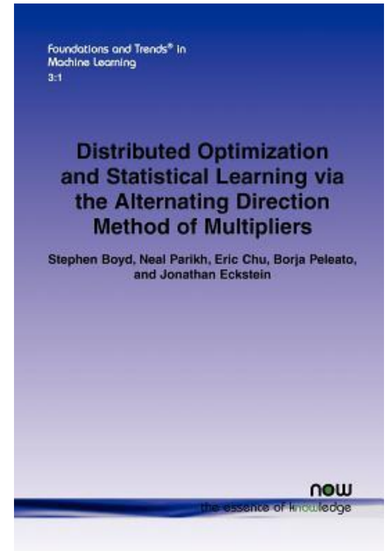


Semidefinite Relaxation (SDR)

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

ADMM

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



Semidefinite Relaxation (SDR)

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

Generic Formulation

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

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

Scalability?

Generic Formulation

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

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

Scalability?

Generic Formulation

	dual vars
$\max \ \langle \boldsymbol{C}, \boldsymbol{X} \rangle$	
s.t. $\mathcal{A}(\boldsymbol{X}) = \boldsymbol{b},$	$\boldsymbol{y}$
$\mathcal{B}(\boldsymbol{X}) \geq \boldsymbol{0},$	$\boldsymbol{z} \geq \boldsymbol{0}$
$\boldsymbol{X} \succeq \boldsymbol{0}.$	$\boldsymbol{S} \succeq \boldsymbol{0}$

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

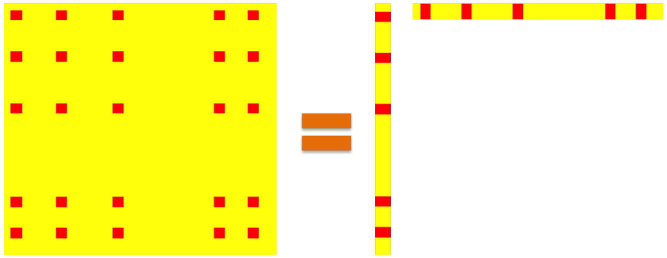
$$\boldsymbol{X}^{(t)} = \underbrace{\left(\boldsymbol{X}^{(t-1)} - \frac{\boldsymbol{C} + \mathcal{A}^{*}\left(\boldsymbol{y}^{(t)}\right) - \mathcal{B}^{*}\left(\boldsymbol{z}^{(t)}\right)}{\mu} \right)_{\succeq \boldsymbol{0}}}_{\text{projection onto PSD cone}}$$

- *Eigen-decomposition of dense matrices is expensive!*

Accelerated ADMM (SDPAD-LR)

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

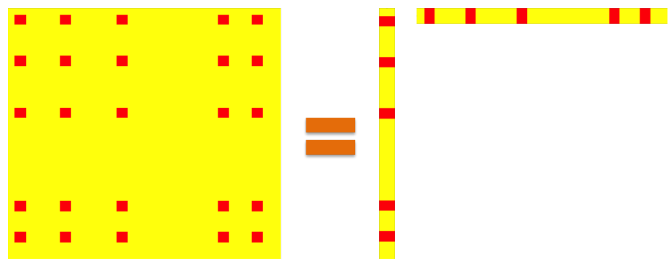
- **Recall: the ground truth obeys rank$(\boldsymbol{X}) = 1$**
  - *Enforce / Exploit Low-Rank Structure!*



Accelerated ADMM (SDPAD-LR)

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

- **Recall: the ground truth obeys rank$(\boldsymbol{X}) = 1$**
  - *Enforce / Exploit Low-Rank Structure!*



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

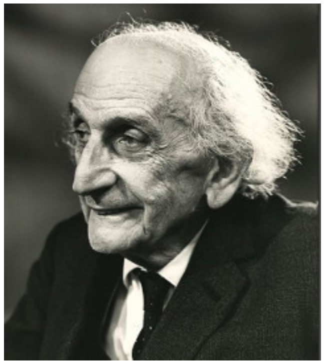
$$\text{eigens of} \quad \left(\underbrace{\boldsymbol{Y}^{(t-1)}\boldsymbol{Y}^{(t-1)\top}}_{\text{low rank}} - \underbrace{\frac{\boldsymbol{C} + \mathcal{A}^*\left(\boldsymbol{y}^{(t)}\right) - \mathcal{B}^*\left(\boldsymbol{z}^{(t)}\right)}{\mu}}_{\text{sparse}} \right)$$

Accelerated ADMM (SDPAD-LR)

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

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

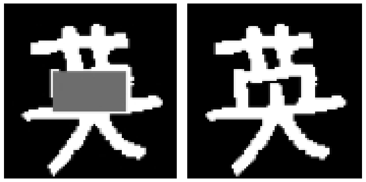
- Numerically fast
  - e.g. **Lanczos Process** $O(nmr^2 + m^2|\mathcal{G}|)$
- Empirically, $r \approx 8$



Cornelius Lanczos

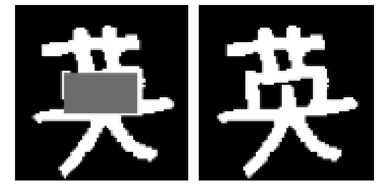
Benchmark Data Sets

- **Benchmark**
  - OPENGM2
  - PIC
  - ORIENT



Benchmark Data Sets

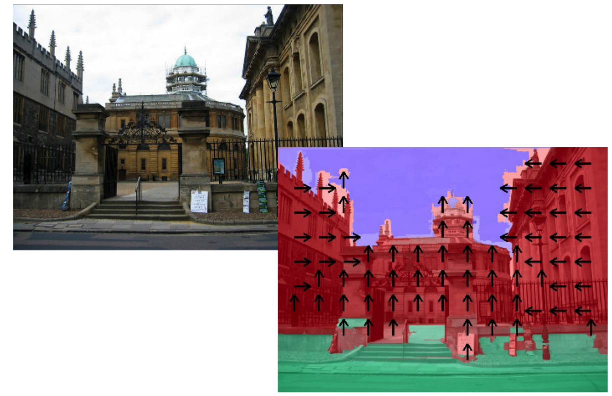
- **Benchmark**
  - OPENGM2
  - PIC
  - ORIENT



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!

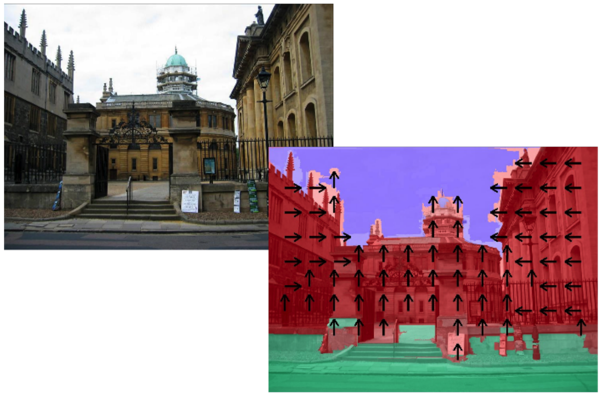
Empirical Convergence: Example



- 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 $< 10^{-3}$



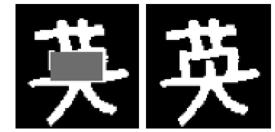
	SDPAD-LR (our algorithm)	SDPAD (original ADMM by Wen'10)
time	21:33	41:33:21
duality gap	5.1×10^{-4}	1.2×10^{-4}
primal-dual infeasibility	1.3×10^{-6}	3.1×10^{-6}

	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!*

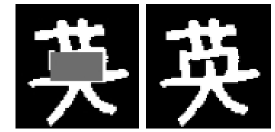
Performance on MAP Problems

- **Performance Measures**
  - mean objective values
  - the percentage of an algorithm achieving the best result



Performance on MAP Problems

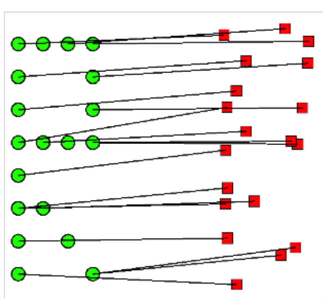
- Performance Measures
  - mean objective values
  - the percentage of an algorithm achieving the best result



	SDPAD-LR	Ficolofo	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	Ficolofo	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%



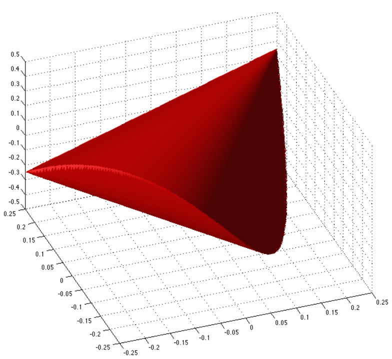
- SDPAD-LR is ***top-performing*** on all datasets

  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?