Constrained Spectral Clustering through Affinity Propagation

Zhengdong Lu — Dept. of CSEE, OGI, OHSU

Miguel Á. Carreira-Perpiñán — EECS, University of California, Merced

1 Constrained Spectral Clustering

Pairwise Constraints:

- **Must-link:** Sample A and B should be in one cluster.
- **Cannot-link:** Sample A and B should NOT be in one cluster.

Constrained Spectral Clustering

We attempt to incorporate the pairwise constraints into affinity matrix, and then apply an off-the-shelf spectral clustering algorithm.

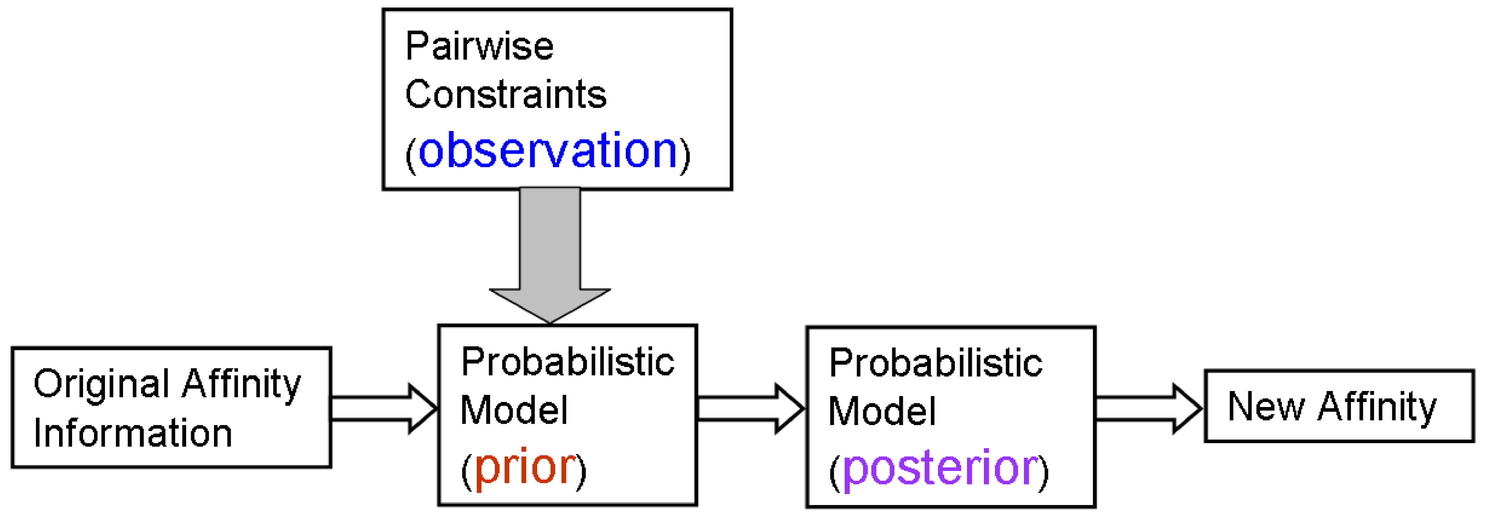
- Pairwise constraints provide another source of affinity information. They are informative, but usually sparse
- If the pairwise constraints are enforced ONLY on constrained sample pairs, its effect on the spectrum of affinity matrix is often marginal.

We want to "propagate" the informative affinity from pairwise constraints to other entries in the affinity matrix.

2 Affinity Propagation (part A): Gaussian Process Interpretation

Key Ideas:

- View the affinity matrix $\mathbf{K}$ (if $\succeq 0$) as the covariance matrix of a zero-mean Gaussian process.
- Express the pairwise constraints as (noisy) observations.



Prior: The original affinity matrix $\mathbf{K} \succ 0$ as the covariance matrix of a zero-mean Gaussian process f:

$$P(\mathbf{f}) = |2\pi\mathbf{K}|^{-N/2} e^{-\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}}$$

where $\mathbf{f} = (f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_N))^T$ at the data points.

Observation: We now treat the given pairwise constraints as a kind of observation, called Ω:

(M) If $\mathbf{x}_i$ and $\mathbf{x}_j$ are must-linked, we assume it is observed that $f(\mathbf{x}_i) - f(\mathbf{x}_j) \sim \mathcal{N}(0, \epsilon_m^2)$

(C) If $\mathbf{x}_i$ and $\mathbf{x}_j$ are cannot-linked, we assume it is observed that $f(\mathbf{x}_i) + f(\mathbf{x}_j) \sim \mathcal{N}(0, \epsilon_c^2)$

where ϵ_m and ϵ_c soften the constraints.

Posterior: From Bayes' rule, the posterior probability of $\mathbf{f}$ given Ω is:

$$P(\mathbf{f}|\Omega) \propto \exp\left(-\frac{1}{2}\mathbf{f}^T\mathbf{K}^{-1}\mathbf{f}\right) \times \exp\left(-\sum_{ij\in\mathcal{M}} \frac{(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2}{2\epsilon_m^2} - \sum_{ij\in\mathcal{C}} \frac{(f(\mathbf{x}_i) + f(\mathbf{x}_j))^2}{2\epsilon_c^2}\right).$$

We propose to use $\overline{\mathbf{K}}_{ij} \equiv \mathrm{E}\{f(\mathbf{x}_i)f(\mathbf{x}_j)|\Omega\}$ as the new affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$.

3 Affinity Propagation (part B): The Algorithm and Examples

We propose to use $\overline{\mathbf{K}}_{ij} \equiv \mathrm{E}\{f(\mathbf{x}_i)f(\mathbf{x}_j)|\Omega\}$ as the new affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$. Since $\mathbf{f}|\Omega$ is still a Gaussian and $\mathrm{E}\{\mathbf{f}|\Omega\} = \mathbf{0}$, we have the following key result:

$$\overline{\mathbf{K}} = (\mathbf{K}^{-1} + \mathbf{M})^{-1} = \mathbf{K} - \mathbf{K}(\mathbf{I} + \mathbf{MK})^{-1}\mathbf{MK} \qquad (1)$$

where $\mathbf{M}$ is a matrix specified by the user
$$\mathbf{M}_{ij} = \begin{cases} \frac{m_i}{\epsilon_m^2} + \frac{c_i}{\epsilon_c^2} & \text{if } i = j \\ -\frac{1}{\epsilon_m^2} & (i,j) \in \mathcal{M} \\ \frac{1}{\epsilon_c^2} & (i,j) \in \mathcal{C} \\ 0 & \text{otherwise} \end{cases}$$

Algorithms

Algorithm A (for two classes) Given $\mathbf{K} \succ 0$:

1. Compose the matrix $\mathbf{M}$ according to Equation (1) based on all constraints and let $\overline{\mathbf{K}} = (\mathbf{K}^{-1} + \mathbf{M})^{-1}$.
2. Let $\mathbf{A}_{ij} = \max(0, \overline{\mathbf{K}}_{ij})\ \forall i, j$.
3. Do spectral clustering with $\mathbf{A}$ as the affinity matrix.

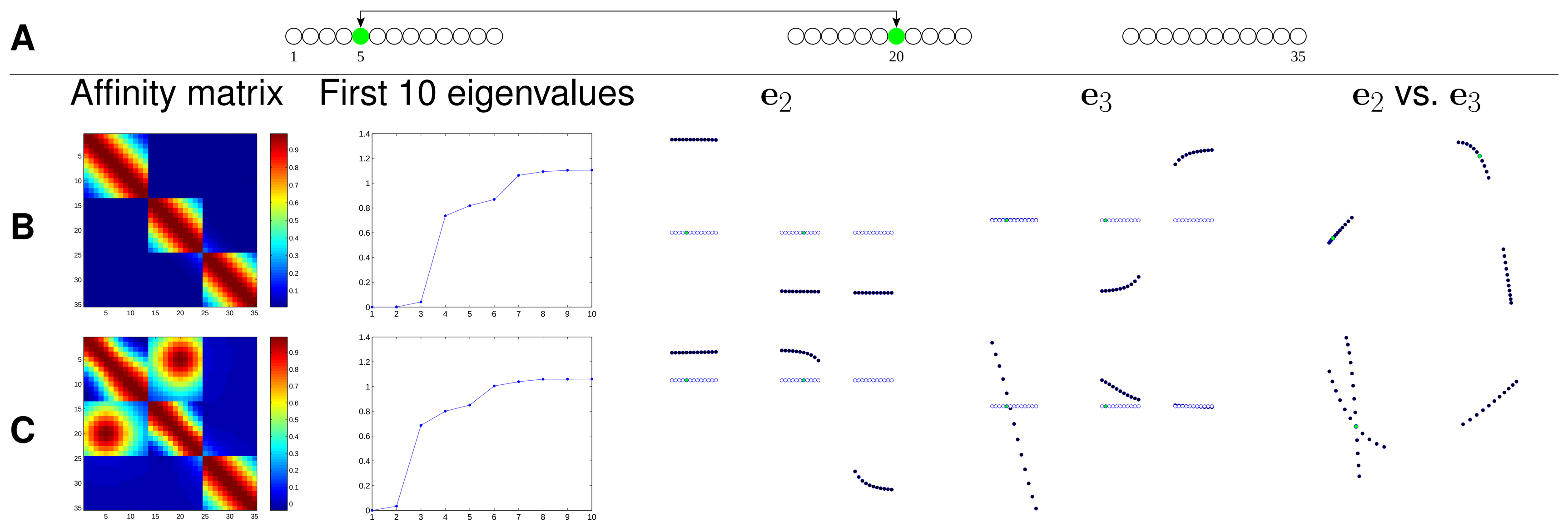
Algorithm B (for more than two classes) Given $\mathbf{K} \succ 0$:

1a. Compose the matrix $\mathbf{M}^m$ according to Equation (1) based on only must-links and let $\mathbf{K}^m = (\mathbf{K}^{-1} + \mathbf{M}^m)^{-1}$.

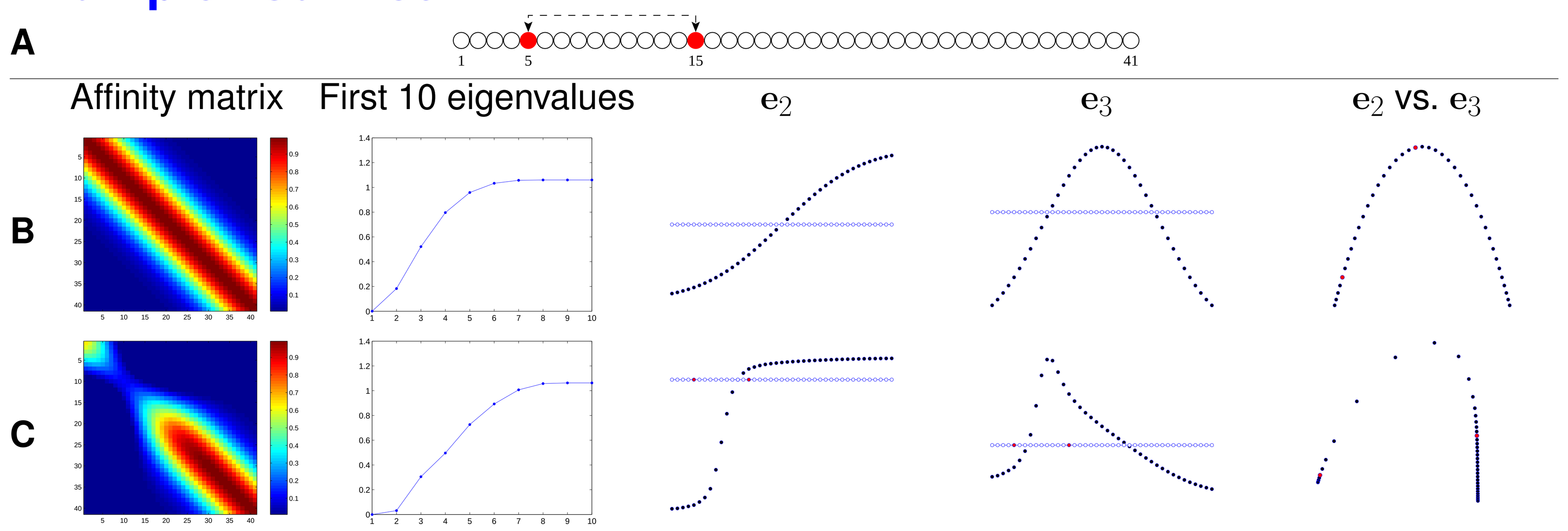
1b. Suppose we have n_c cannot-links. Compose the matrix $\mathbf{M}^{c,k}, k = \{1, 2, \dots, n_c\}$ according to Equation (1) based on the i^{th} cannot-link and let $\mathbf{K}^{c,k} = ((\mathbf{K}^m)^{-1} + \mathbf{M}^{c,k})^{-1}$.

2. Let $\mathbf{A}_{ij} = \max\left(0, \min(\mathbf{K}_{ij}^{c,1}, \dots, \mathbf{K}_{ij}^{c,n_c})\right)\ \forall i, j$.
3. Do spectral clustering with $\mathbf{A}$ as the affinity matrix.

Example: Must-link



Example: Cannot-link



4 Minimum-Divergence Formulation

The posterior probability of $\mathbf{f}|\Omega$ can be equivalently formulated as a minimum KL-divergence problem

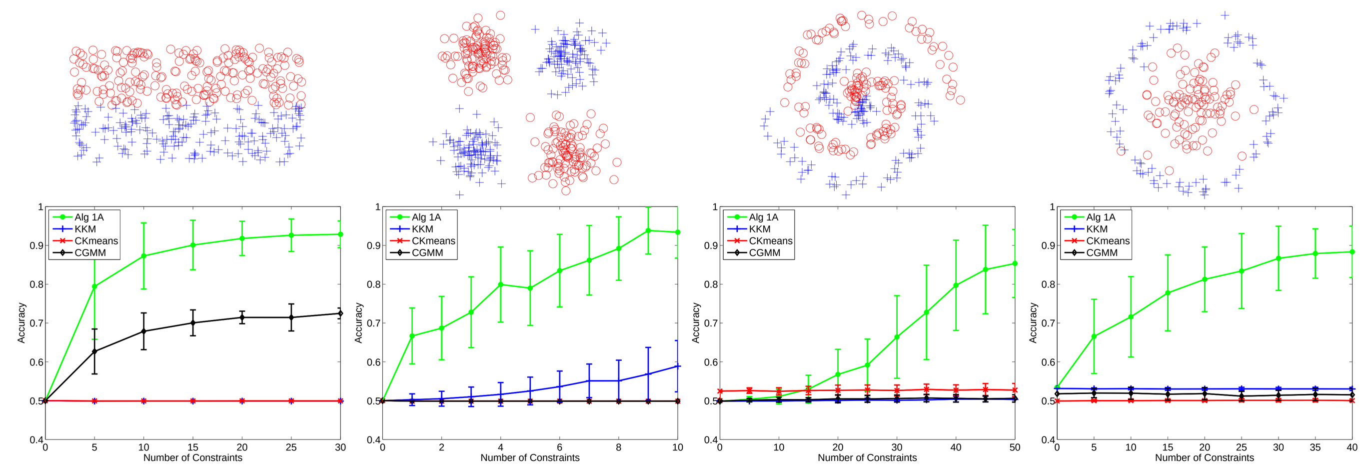
$$\min_{\overline{P}} \ \mathrm{D}\left(\overline{P}(\mathbf{f}) \| P(\mathbf{f})\right)$$
$$\text{s.t.} \quad \mathrm{E}_{\overline{P}}\left\{(f(\mathbf{x}_i) - f(\mathbf{x}_j))^2\right\} \le \alpha_{ij}, \quad (i,j) \in \mathcal{M}$$
$$\mathrm{E}_{\overline{P}}\left\{(f(\mathbf{x}_i) + f(\mathbf{x}_j))^2\right\} \le \beta_{ij}, \quad (i,j) \in \mathcal{C}$$

which leads to the following matrix nearness problem

$$\min_{\overline{\mathbf{K}}} \ \log\left(|\mathbf{K}| / |\overline{\mathbf{K}}|\right) + \mathrm{tr}\left(\mathbf{K}^{-1}\overline{\mathbf{K}}\right)$$
$$\text{s.t.} \quad \overline{\mathbf{K}} \succ 0$$
$$\overline{\mathbf{K}}_{ii} + \overline{\mathbf{K}}_{jj} - 2\overline{\mathbf{K}}_{ij} \le \alpha_{ij}, \quad (i,j) \in \mathcal{M}$$
$$\overline{\mathbf{K}}_{ii} + \overline{\mathbf{K}}_{jj} + 2\overline{\mathbf{K}}_{ij} \le \beta_{ij}, \quad (i,j) \in \mathcal{C}.$$

5 Experiments

On Artificial Data Sets



Semi-supervised Image Segmentation