digiteo

Recherche en sciences & technologies de l'information

FORUM 2011

Ecole Polytechnique

18 octobre 2011

Toby Dylan Hocking, Armand Joulin, Francis Bach, and Jean-Philippe Vert

Nº 6

Clusterpath: an algorithm for clustering using convex fusion penalties

1 The clustering problem

Clustering: label n points in p dimensions $X \in \mathbb{R}^{n \times p}$.

- $X_i \in \mathbb{R}^p$ row/datum i of X.
- $\alpha_i \in \mathbb{R}^p$ clustered X_i.
- $\alpha^j \in \mathbb{R}^n$ column/variable j.

Methods:

- K-means
- Hierarchical
- Mixture models
- Spectral (Ng *et al.* 2001)

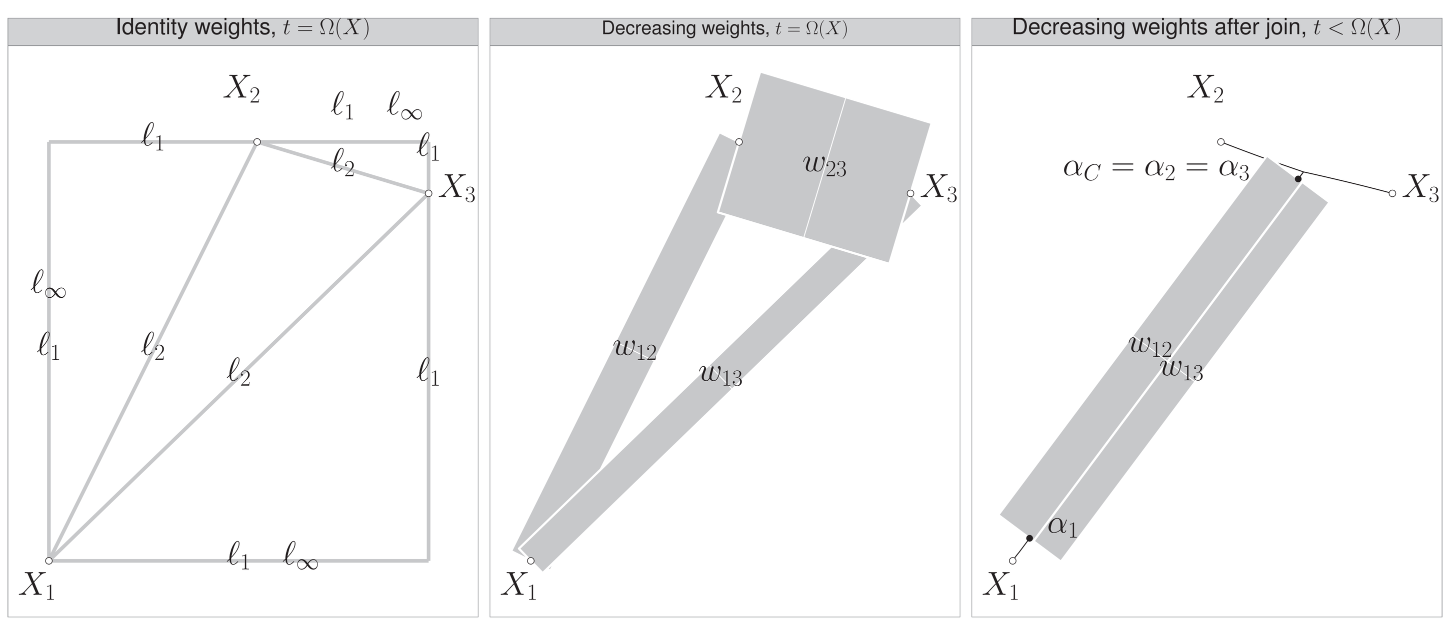
Issues:

- Hierarchy
- Convexity
- Greediness
- Stability
- Interpretability

2 Clusterpath: relaxing a fusion penalty

- Hard-thresholding of differences is a combinatorial problem: $\min_{\alpha \in \mathbb{R}^{n \times p}} ||\alpha - X||_F^2$ subject to $\sum_{i<j} 1_{\alpha_i \neq \alpha_j} \leq t$
- Relaxation: $\Omega(\alpha) = \sum_{i<j} ||\alpha_i - \alpha_j||_q w_{ij} \leq t, \quad w_{ij} = \exp(-\gamma ||X_i - X_j||_2^2)$

Geometric interpretation: constrain area between points

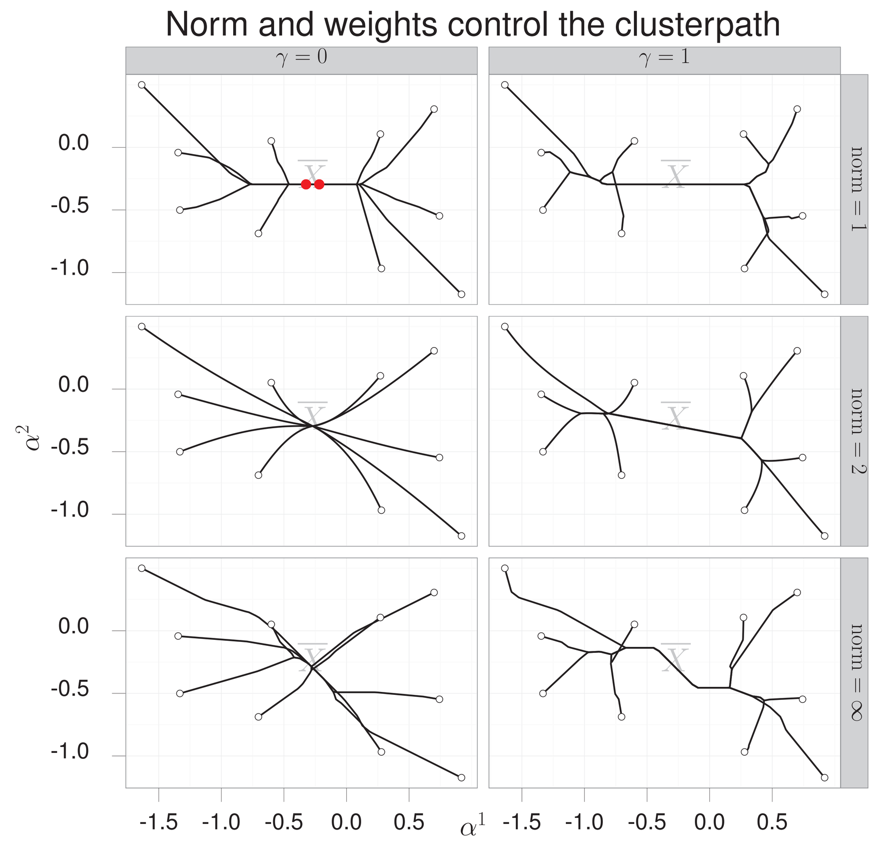


- The Lagrange form is useful for optimization algorithms:

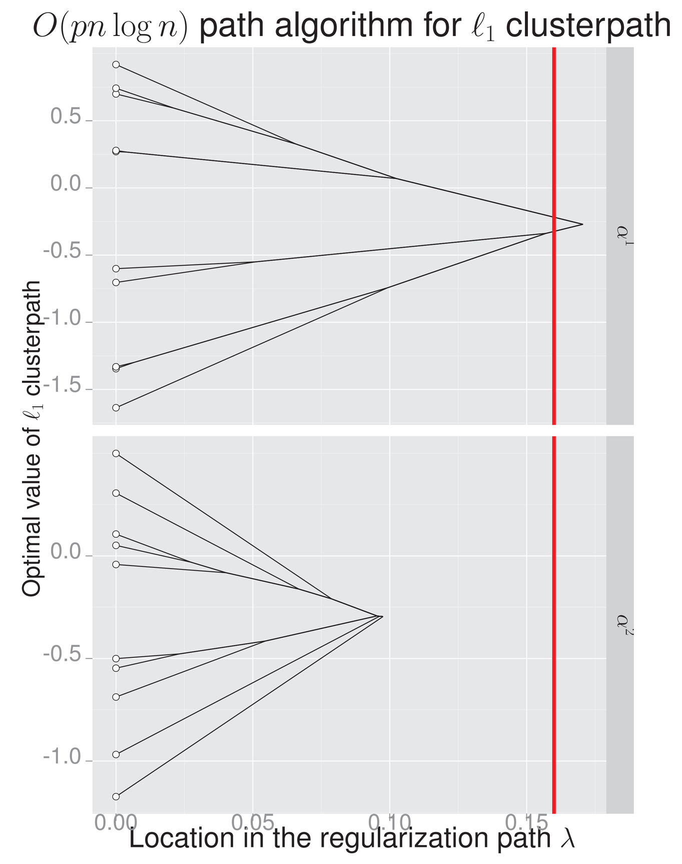
$$\alpha^*(X, \lambda, q, w) = \operatorname{argmin}_{\alpha \in \mathbb{R}^{n \times p}} \frac{1}{2} ||\alpha - X||_F^2 + \lambda \sum_{i<j} ||\alpha_i - \alpha_j||_q w_{ij}$$

- The **clusterpath** of X is the path of optimal α^* obtained by varying λ, for fixed weights $w_{ij} > 0$ and norm $q \in \{1, 2, \infty\}$.

Norm and weights control the clusterpath



$O(pn \log n)$ path algorithm for ℓ_1 clusterpath



Related work: "Fused lasso" Tibshirani and Saunders (2005), "grouping pursuit" Shen and Huang (2010), "sum of norms" Lindsten *et al.* (2011).

3 Outline of the ℓ_1 path algorithm

Consider one variable: $\alpha, X \in \mathbb{R}^n$. Condition sufficient for optimality:

$$0 = \alpha_i - X_i + \lambda \sum_{\substack{j \neq i \\ \alpha_i \neq \alpha_j}} w_{ij} \operatorname{sign}(\alpha_i - \alpha_j) + \lambda \sum_{\substack{j \neq i \\ \alpha_i = \alpha_j}} w_{ij} \beta_{ij},$$

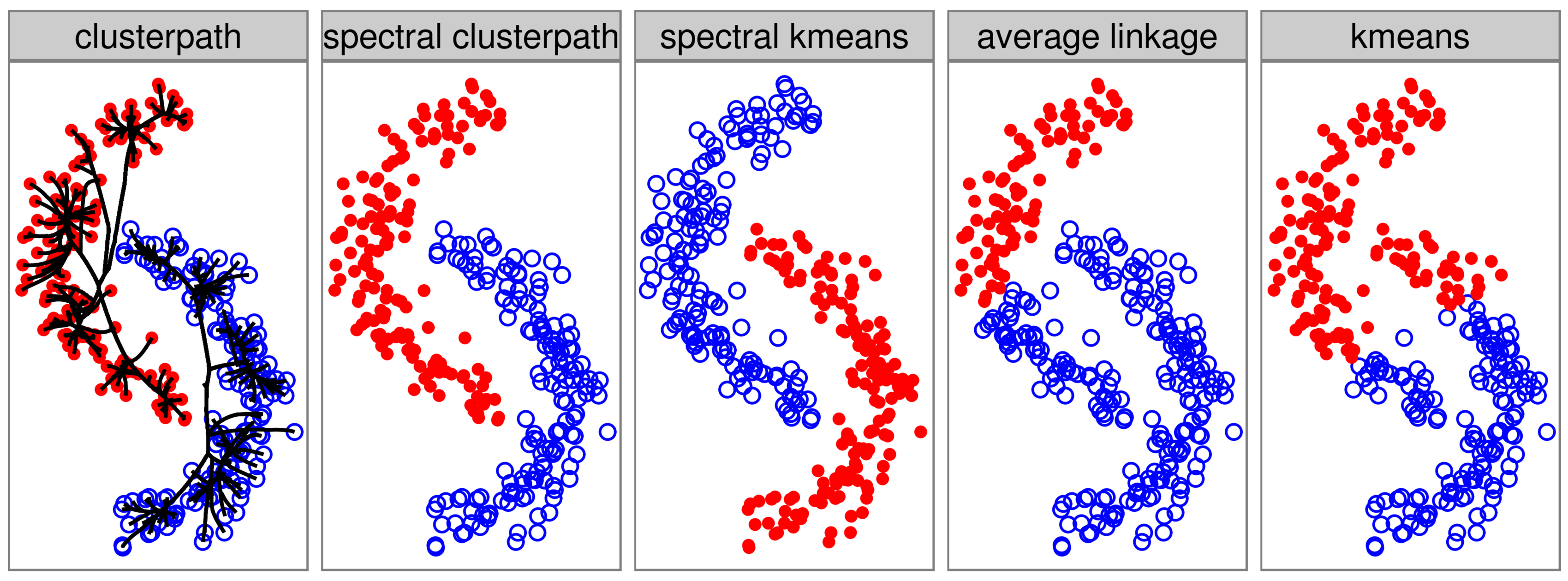
with $|\beta_{ij}| \leq 1$ and $\beta_{ij} = -\beta_{ji}$ (Hoefling 2009).

1. For $\lambda = 0$ the solution $\alpha = X$ is optimal. Initialize clusters $C_i = \{i\}$.
2. As λ increases, the solutions will follow straight lines:

$$\frac{d\alpha_C}{d\lambda} = v_C = \sum_{j \notin C} w_{jC} \operatorname{sign}(\alpha_j - \alpha_C) = \sum_{j \notin C} \sum_{i \in C} w_{ij} \operatorname{sign}(\alpha_j - \alpha_C)$$

3. When 2 clusters C_1 and C_2 fuse, they form a new cluster $C = C_1 \cup C_2$ with $v_C = (|C_1|v_1 + |C_2|v_2)/(|C_1| + |C_2|)$.
4. Stop when all the points merge at the mean $\bar{X}$.

4 Performs similarly to spectral clustering



5 Free software implementation

- `http://clusterpath.r-forge.r-project.org/`
- Dedicated C++ optimization algorithms with R interface.
  - Calculates the exact ℓ_1 clusterpath for identity weights.
  - Active-set algorithm for $\ell_{1|2}$ clusterpath with general weights.
- R interface to Python `cvxmod` clusterpath solver.

6 Future work

- Necessary and sufficient conditions for cluster splitting?
- Consistence of the clusterpath tree? (Hartigan, 1975)
- Automatically learning weights and number of clusters?
- Applications to solving proximal problems.

îledeFrance

Contact : Toby.Hocking@inria.fr

inria informatics mathematics