Probably certifiably correct k-means clustering

Dustin G. Mixon

AIR FORCE INSTITUTE OF TECHNOLOGY

Compressed Sensing and its Applications

December 7 – 11, 2015

This week in clustering

This week in clustering

Regularize graphs before spectral clustering

This week in clustering

Regularize graphs before spectral clustering

k-means SDP clusters random data

This week in clustering

Regularize graphs before spectral clustering

k-means SDP clusters random data

SDP beats spectral clustering (and MLE!)

This week in clustering

Regularize graphs before spectral clustering

k-means SDP clusters random data

SDP beats spectral clustering (and MLE!)

This talk: Use SDP to quickly certify optimal clusterings

Collaborators

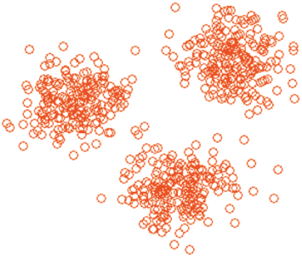
Takayuki Iguchi
AFIT

Jesse Peterson
AFIT

Soledad Villar
UT Austin

The k-means problem

Given a point cloud, partition the points into concentrated clusters



k-means objective:

$$\sum_{t=1}^{k} \sum_{i \in C_t} \left\| x_i - \frac{1}{|C_t|} \sum_{j \in C_t} x_j \right\|^2$$

- NP-hard to minimize in general
- Lloyd's algorithm often works, but no optimality certificate

An SDP relaxation

Taking $D_{ij} := \|x_i - x_j\|^2$, then

$$\sum_{t=1}^{k} \sum_{i \in C_t} \left\| x_i - \frac{1}{|C_t|} \sum_{j \in C_t} x_j \right\|^2 = \frac{1}{2} \operatorname{Tr}\left(D \sum_{t=1}^{k} \frac{1}{|C_t|} 1_{C_t} 1_{C_t}^\top \right)$$

Peng, Wei, SIAM J. Optim., 2007

An SDP relaxation

Taking $D_{ij} := \|x_i - x_j\|^2$, then

$$\sum_{t=1}^{k} \sum_{i \in C_t} \left\| x_i - \frac{1}{|C_t|} \sum_{j \in C_t} x_j \right\|^2 = \frac{1}{2} \operatorname{Tr}\left(D \sum_{t=1}^{k} \frac{1}{|C_t|} 1_{C_t} 1_{C_t}^\top \right)$$

Proof:

$$\begin{aligned}
\operatorname{Tr}(D 1_{C_t} 1_{C_t}^\top) &= \sum_{i \in C_t} \sum_{j \in C_t} \|x_i - x_j\|^2 \\
&= \sum_{i \in C_t} \sum_{j \in C_t} \left\| (x_i - c_t) - (x_j - c_t) \right\|^2 \\
&= 2|C_t| \sum_{i \in C_t} \|x_i - c_t\|^2
\end{aligned}$$

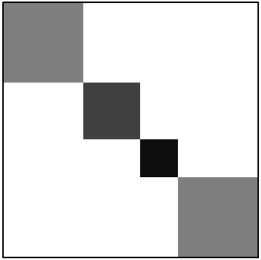
Divide by $2|C_t|$ and add. $\square$

Peng, Wei, SIAM J. Optim., 2007

An SDP relaxation

Taking $D_{ij} := \|x_i - x_j\|^2$, then

$$\sum_{t=1}^{k} \sum_{i \in C_t} \left\| x_i - \frac{1}{|C_t|} \sum_{j \in C_t} x_j \right\|^2 = \frac{1}{2} \operatorname{Tr} \left(D \underbrace{\sum_{t=1}^{k} \frac{1}{|C_t|} 1_{C_t} 1_{C_t}^\top} \right)$$

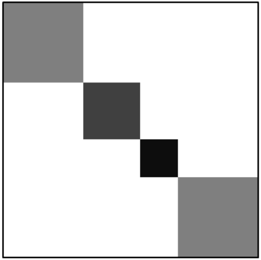


Peng, Wei, SIAM J. Optim., 2007

An SDP relaxation

Taking $D_{ij} := \|x_i - x_j\|^2$, then

$$\sum_{t=1}^{k} \sum_{i \in C_t} \left\| x_i - \frac{1}{|C_t|} \sum_{j \in C_t} x_j \right\|^2 = \frac{1}{2} \operatorname{Tr} \left(D \underbrace{\color{red}{\sum_{t=1}^{k} \frac{1}{|C_t|} 1_{C_t} 1_{C_t}^{\top}}} \right)$$



Relax to SDP:

$$\begin{array}{ll} \text{minimize} & \operatorname{Tr}(D\color{red}{X}) \\ \text{subject to} & \operatorname{Tr}(X) = k \\ & X1 = 1 \\ & X \geq 0 \\ & X \succeq 0 \end{array}$$

Peng, Wei, SIAM J. Optim., 2007

Main problem

SDP solvers are polytime, but slow

SDP clusters 64 points in 20 sec, Lloyd takes 0.001 sec

(cf. PhaseLift vs. Wirtinger flow)

Main problem

SDP solvers are polytime, but slow

SDP clusters 64 points in 20 sec, Lloyd takes 0.001 sec

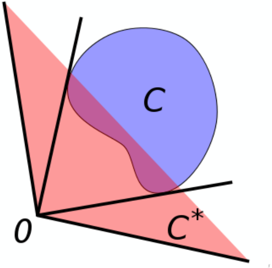
(cf. PhaseLift vs. Wirtinger flow)

Probably certifiably correct algorithm

- Oracle provides k-means-optimal solution whp
- Task: Given "solution," quickly certify optimality

Preliminaries

Dual cone: $C^* := \{x : \langle x, y\rangle \geq 0 \ \ \forall y \in C\}$



Preliminaries

Dual cone: $C^* := \{x : \langle x, y \rangle \geq 0 \ \ \forall y \in C\}$

Primal program:

$$\begin{aligned} \max \quad & \langle c, x \rangle \\ \text{s.t.} \quad & b - Ax \in L \\ & x \in K \end{aligned}$$

Dual program:

$$\begin{aligned} \min \quad & \langle b, y \rangle \\ \text{s.t.} \quad & A^\top y - c \in K^* \\ & y \in L^* \end{aligned}$$

Preliminaries

Dual cone: $C^* := \{x : \langle x, y\rangle \geq 0 \quad \forall y \in C\}$

Primal program:

$$\begin{aligned} \max \quad & \langle c, x\rangle \\ \text{s.t.} \quad & b - Ax \in L \\ & x \in K \end{aligned}$$

Dual program:

$$\begin{aligned} \min \quad & \langle b, y\rangle \\ \text{s.t.} \quad & A^\top y - c \in K^* \\ & y \in L^* \end{aligned}$$

Weak duality: $\langle b - Ax, y\rangle \geq 0$, $\langle x, A^\top y - c\rangle \geq 0$

$$\implies \qquad \langle c, x\rangle \leq \langle x, A^\top y\rangle = \langle Ax, y\rangle \leq \langle b, y\rangle$$

Strong duality: $\langle c, x_{\mathrm{opt}}\rangle = \langle b, y_{\mathrm{opt}}\rangle$ "dual certificate"

Preliminaries

Dual cone: $C^* := \{x : \langle x, y\rangle \geq 0 \;\; \forall y \in C\}$

Primal program:

$$\begin{aligned} \max \quad & \langle c, x\rangle \\ \text{s.t.} \quad & b - Ax \in L \\ & x \in K \end{aligned}$$

Dual program:

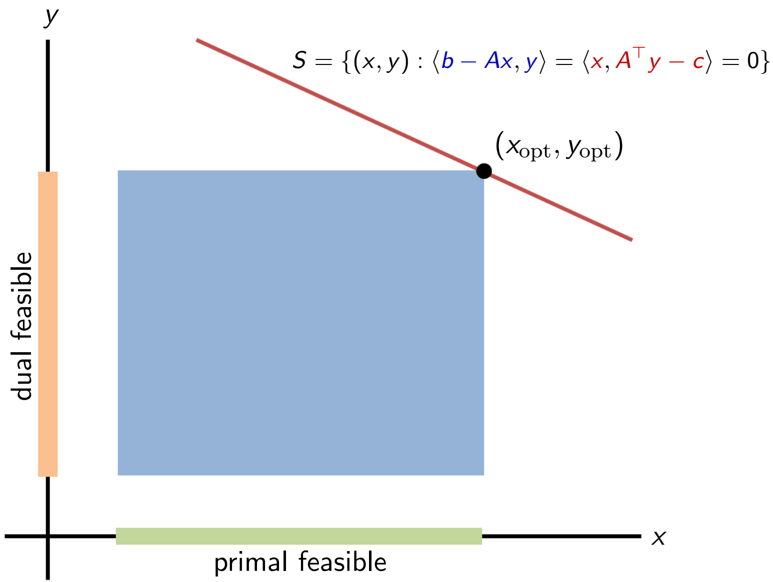
$$\begin{aligned} \min \quad & \langle b, y\rangle \\ \text{s.t.} \quad & A^\top y - c \in K^* \\ & y \in L^* \end{aligned}$$

Complementary slackness

x is primal-opt and y is dual-opt if and only if

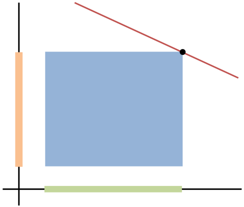
- x is primal feasible
- y is dual feasible
- $\langle b - Ax, y\rangle = \langle x, A^\top y - c\rangle = 0$

Preliminaries



The big idea (Afonso Bandeira)

Task: Given x_{opt}, quickly find y_{opt}



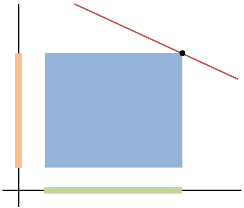
Method:

1. Check that x_{opt} is primal feasible
2. Find y such that $(x_{\mathrm{opt}}, y) \in S$
3. Check that y is dual feasible

Bandeira, arXiv:1509.00824

The big idea (Afonso Bandeira)

Task: Given x_{opt}, quickly find y_{opt}



Method:

1. Check that x_{opt} is primal feasible
2. Find y such that $(x_{\mathrm{opt}}, y) \in S$
3. Check that y is dual feasible

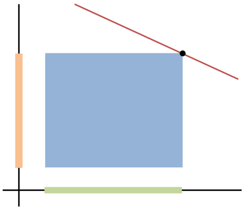
Example: Minimum bisection in stochastic block model

- Easy to find **unique** y such that $(x_{\mathrm{opt}}, y) \in S$
- Checking dual feasibility is an eigenvalue problem (easy)

Bandeira, `arXiv:1509.00824`

The big idea (Afonso Bandeira)

Task: Given x_{opt}, quickly find y_{opt}



Method:

1. Check that x_{opt} is primal feasible
2. Find y such that $(x_{\text{opt}}, y) \in S$
3. Check that y is dual feasible

Example: Minimum bisection in stochastic block model

- Easy to find **unique** y such that $(x_{\text{opt}}, y) \in S$
- Checking dual feasibility is an eigenvalue problem (easy)

Problem: y is **not unique** in the case of k-means
(Choice of y is an art form, "optimal" choice remains open)

Bandeira, arXiv:1509.00824

A small technicality

Subproblem in checking dual feasibility:

Is $\operatorname{span}(v)$ **the unique leading eigenspace of** A**?**

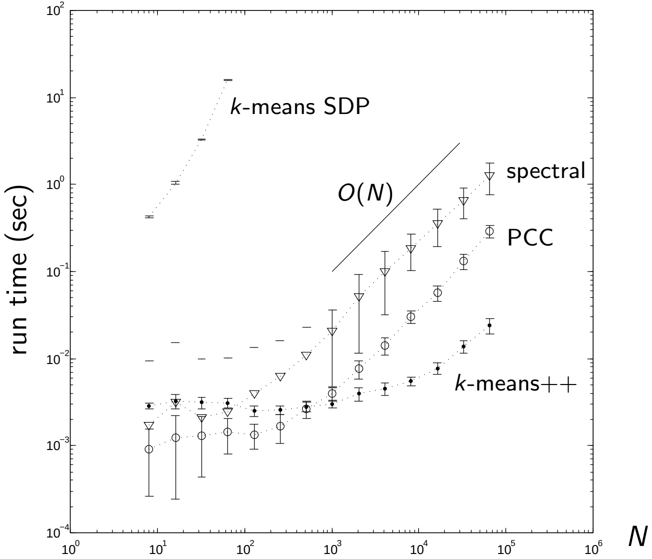
Fast solution: Power method from random initialization

Report $1-\eta$ confidence after $O(\log(1/\eta))$ power iterations

Open problem: Remove the possibility of "false certificates"

Iguchi, M., Peterson, Villar, `arXiv:1509.07983`

It works, and it's fast!



Iguchi, M., Peterson, Villar, `arXiv:1509.07983`

Guarantee for random problem instances

$(\mathcal{D}, \gamma, n)$**-stochastic ball model**

- $\mathcal{D}$ = rotation-invariant distribution over unit ball in $\mathbb{R}^m$
- $\gamma_1, \ldots, \gamma_k$ = ball centers in $\mathbb{R}^m$
- Draw $r_{t,1}, \ldots, r_{t,n}$ i.i.d. from $\mathcal{D}$ for each $i \in \{1, \ldots, k\}$
- $x_{t,i} = \gamma_t + r_{t,i}$ = ith point from cluster t

When does the PCC method certify the planted solution whp?

Nellore, Ward, arXiv:1309.3256

Iguchi, M., Peterson, Villar, arXiv:1509.07983

Guarantee for random problem instances

$(\mathcal{D}, \gamma, n)$**-stochastic ball model**

- $\mathcal{D}$ = rotation-invariant distribution over unit ball in $\mathbb{R}^m$
- $\gamma_1, \ldots, \gamma_k$ = ball centers in $\mathbb{R}^m$
- Draw $r_{t,1}, \ldots, r_{t,n}$ i.i.d. from $\mathcal{D}$ for each $i \in \{1, \ldots, k\}$
- $x_{t,i} = \gamma_t + r_{t,i}$ = ith point from cluster t

When does the PCC method certify the planted solution whp?

Theorem

PCC certifies the planted solution under $(\mathcal{D}, \gamma, n)$-SBM w.p. $1 - e^{-\Omega_{\mathcal{D},\gamma}(n)}$ if

$$\min_{i \neq j} \|\gamma_i - \gamma_j\| \geq 2 + \frac{k^2}{m}$$

Nellore, Ward, arXiv:1309.3256

Iguchi, M., Peterson, Villar, arXiv:1509.07983

When does SDP recover planted clustering?

Corollary

SDP recovers the planted solution under $(\mathcal{D}, \gamma, n)$-SBM w.p. $1 - e^{-\Omega_{\mathcal{D},\gamma}(n)}$ if

$$\min_{i \neq j} \|\gamma_i - \gamma_j\| \geq \min \left\{ 2 + \frac{k^2}{m},\ 2\sqrt{2}(1 + \frac{1}{\sqrt{m}}) \right\}$$

Bounds from different choices of dual certificate (art form)

Appears loose in the small-m regime

What is the best bound? (Ideas from statistical mechanics?)

Awasthi, Bandeira, Charikar, Krishnaswamy, Villar, Ward, Proc. ITCS, 2015

Iguchi, M., Peterson, Villar, arXiv:1509.07983

When does SDP recover planted clustering?

Natural conjecture: SDP recovers whp provided $\min_{i \neq j} \|\gamma_i - \gamma_j\| > 2$

Iguchi, M., Peterson, Villar, `arXiv:1509.07983`

When does SDP recover planted clustering?

Natural conjecture: SDP recovers whp provided $\min_{i \neq j} \|\gamma_i - \gamma_j\| > 2$

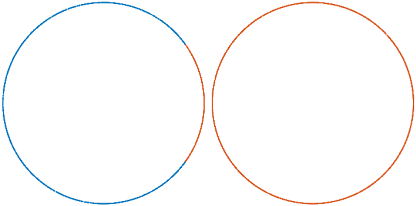
Disproof: Cluster two unit circles in $\mathbb{R}^2$ with $\|\gamma_1 - \gamma_2\| = 2.08$

Iguchi, M., Peterson, Villar, arXiv:1509.07983

When does SDP recover planted clustering?

Natural conjecture: SDP recovers whp provided $\min_{i \neq j} \|\gamma_i - \gamma_j\| > 2$

Disproof: Cluster two unit circles in $\mathbb{R}^2$ with $\|\gamma_1 - \gamma_2\| = 2.08$



The planted clustering is not k-means-optimal!

Open problem: Necessary separation for two $(m-1)$-spheres?

Iguchi, M., Peterson, Villar, arXiv:1509.07983

Conclusion and future directions

Relaxations offer fast optimality certificates

Conclusion and future directions

Relaxations offer fast optimality certificates

Phase retrieval

- Injectivity is hard to check
- Fast uniqueness certificate? Golfing scheme?

PCC approximation ratios

- In some applications, relaxations aren't tight but close
- Stable version of the complementary slackness trick?

Questions?

Probably certifiably correct k-means clustering
T. Iguchi, D. G. Mixon, J. Peterson, S. Villar
arXiv:1509.07983

A note of probably certifiably correct algorithms
A. S. Bandeira
arXiv:1509.00824

Also, google **short fat matrices** for my research blog