

# Sparse PCA with Multiple Components

Ryan Cory-Wright, Jean Pauphilet

# Principal Component Analysis (PCA)

## Idea

For a high-dimensional dataset  $X$ , find a few orthogonal directions ( $r$ ) that explain most of the variance

## Algorithm

Dataset  $X \in \mathbb{R}^{n \times p}$

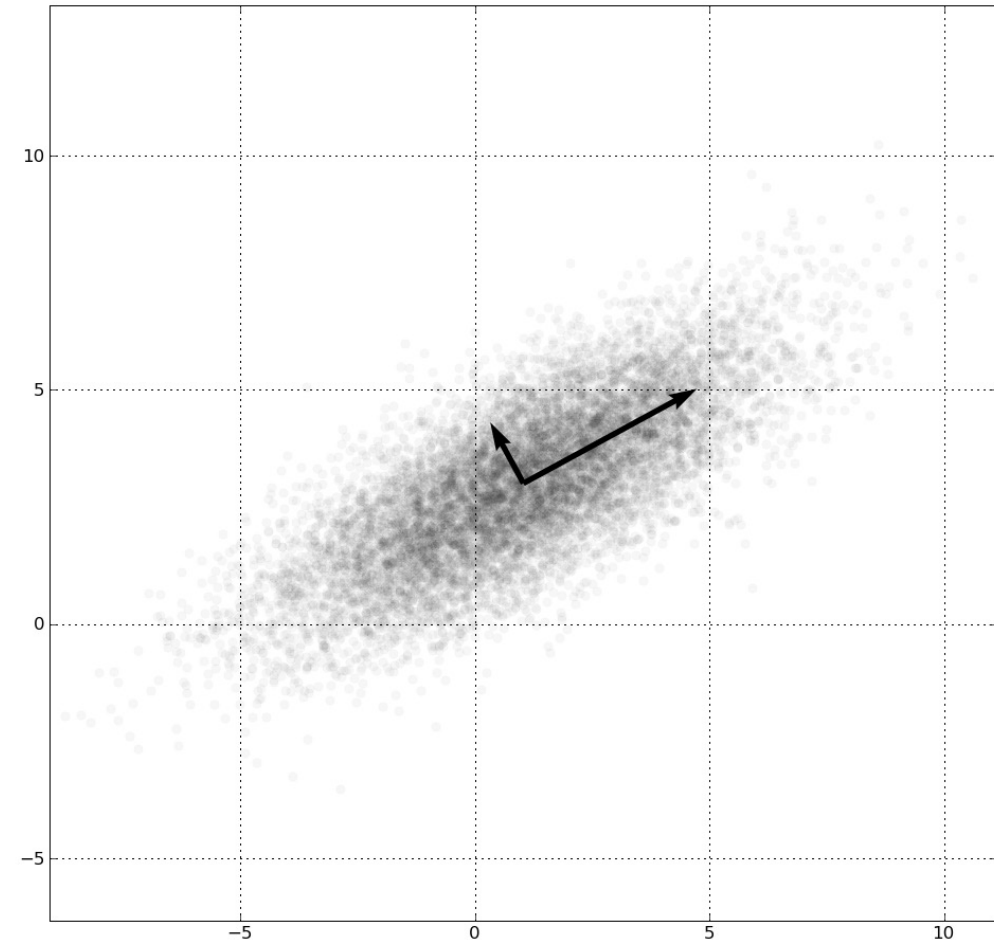
Compute its covariance/correlation matrix  $\Sigma$

Solve

$$\max_{U \in \mathbb{R}^{p \times r}} \langle UU^\top, \Sigma \rangle \text{ s.t. } U^\top U = \mathbb{I}$$

Compress dataset into smaller one

$$X_{\text{comp}} := U_{[1:r]}X \in \mathbb{R}^{n \times r}$$



# Optimization for PCA

$$\max_{\mathbf{U} \in \mathbb{R}^{p \times r}} \langle \mathbf{U}\mathbf{U}^\top, \mathbf{\Sigma} \rangle \text{ s.t. } \mathbf{U}^\top \mathbf{U} = \mathbb{I}$$

It looks hard...

- Maximizing a convex function
- Non-convex quadratic constraints (orthogonality)

... can be solved very efficiently

- Solution obtained via truncated SVD
- For  $r=1$ , orthogonality constraint is trivial to satisfy (scaling)
- For  $r > 1$ , greedy is optimal  
i.e., solve for  $r=1$ , deflate the covariance matrix, repeat  
**Deflation naturally takes care of orthogonality!**

# Why sparsity in PCA?

- **Interpretability**  
New coordinates (PCs) can be dense combination of all features
- **Consistency** in high-dimensional settings  
When  $p/n \rightarrow \alpha$ , PCA can be inconsistent

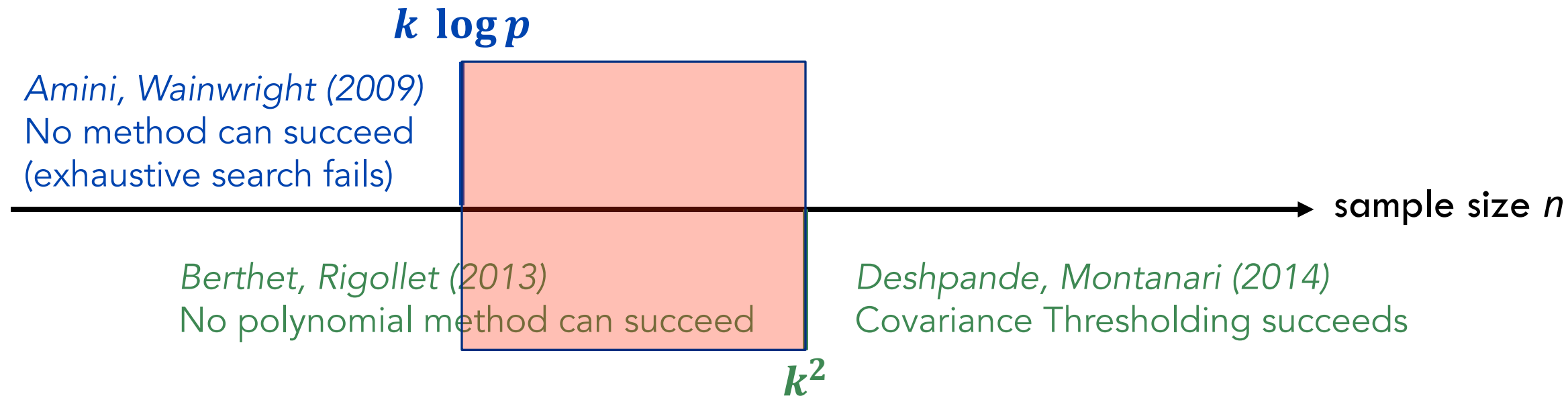
## Solution Sparsity

$$\max_{\mathbf{u} \in \mathbb{R}^p} \langle \mathbf{\Sigma}, \mathbf{u}\mathbf{u}^\top \rangle \text{ s.t. } \|\mathbf{u}\|_2^2 = 1, \|\mathbf{u}\|_0 \leq k$$

PCA requires in the order of  $n \gtrsim p$  samples

sparse PCA requires  $n \gtrsim k \log p$

# Why “provably optimal” algorithms can help?



Opportunity for optimization!

## From $r=1$ to $r>1$

Jungle of algorithms for sparse PCA with  $r=1$  PC

However, for  $r > 1$ , ...

- Deflation approach is no longer optimal
- Deflation no longer guarantees orthogonality (feasibility)
- Assuming perfect support detection, estimation is no longer trivial

# Sparse PCA With Multiple Components

**Overall problem** Explain dataset using **sparse** and **mutually orthogonal** components

$$\max_{\mathbf{U} \in \mathbb{R}^{p \times r}} \langle \mathbf{U}\mathbf{U}^\top, \mathbf{\Sigma} \rangle \quad \text{s.t.} \quad \mathbf{U}^\top \mathbf{U} = \mathbb{I}, \quad \|\mathbf{U}\|_0 \leq k.$$

**OR**

$$\|\mathbf{U}_t\|_0 \leq k_t, \quad \forall t \in [r]$$

## Our Contributions

1. Reformulation as as rank and sparsity constrained problem
2. Tight and scalable semidefinite relaxation w. strong valid inequalities
3. Coupled w. good heuristics, provably near-optimal solutions for  $p \approx 100$

# Reformulation

$$\max_{\mathbf{U} \in \mathbb{R}^{p \times r}} \langle \mathbf{U}\mathbf{U}^\top, \boldsymbol{\Sigma} \rangle \quad \text{s.t.} \quad \mathbf{U}^\top \mathbf{U} = \mathbb{I}, \quad \|\mathbf{U}\|_0 \leq k.$$

$$\max_{\substack{\mathbf{Z} \in \{0,1\}^{p \times r} \\ \langle \mathbf{E}, \mathbf{Z} \rangle \leq k}} \max_{\mathbf{U} \in \mathbb{R}^{p \times r}} \langle \mathbf{U}\mathbf{U}^\top, \boldsymbol{\Sigma} \rangle \quad \text{s.t.} \quad \mathbf{U}^\top \mathbf{U} = \mathbb{I}, \quad U_{i,t} = 0 \text{ if } Z_{i,t} = 0, \quad \forall i \in [p], \forall t \in [r],$$

Introduce

$$\mathbf{Y}^t := \mathbf{U}_t \mathbf{U}_t^\top, \quad \mathbf{Y} := \sum_{t \in [r]} \mathbf{Y}^t = \mathbf{U}\mathbf{U}^\top$$

$$\langle \mathbf{U}\mathbf{U}^\top, \boldsymbol{\Sigma} \rangle \rightarrow \langle \mathbf{Y}, \boldsymbol{\Sigma} \rangle$$

$$\mathbf{U}^\top \mathbf{U} = \mathbb{I} \rightarrow \text{tr}(\mathbf{Y}^t) = 1, \quad \langle \mathbf{Y}^t, \mathbf{Y}^{t'} \rangle = 0 \quad (t \neq t')$$

$$U_{i,t} = 0 \text{ if } Z_{i,t} = 0 \rightarrow Y_{i,j}^t = 0 \text{ if } Z_{i,t} = 0$$

$$\max_{\substack{\mathbf{Z} \in \{0,1\}^{p \times r} \\ \langle \mathbf{E}, \mathbf{Z} \rangle \leq k}} \max_{\mathbf{Y} \in \mathcal{S}^p, \mathbf{Y}^t \in \mathcal{S}_+^p} \langle \mathbf{Y}, \boldsymbol{\Sigma} \rangle \quad \text{s.t.} \quad \text{tr}(\mathbf{Y}^t) = 1, \quad \forall t \in [r], \quad \langle \mathbf{Y}^t, \mathbf{Y}^{t'} \rangle = 0, \quad \forall t, t' \in [r], t \neq t',$$

$$Y_{i,j}^t = 0 \text{ if } Z_{i,t} = 0, \quad \forall t \in [r], i, j \in [p]$$

$$\mathbf{Y} = \sum_{t=1}^r \mathbf{Y}^t, \quad \text{Rank}(\mathbf{Y}^t) = 1, \quad \forall t \in [r].$$



# Reformulation

$$\max_{\substack{\mathbf{Z} \in \{0,1\}^{p \times r}: \\ \langle \mathbf{E}, \mathbf{Z} \rangle \leq k}} \max_{\mathbf{Y} \in \mathcal{S}^p, \mathbf{Y}^t \in \mathcal{S}_+^p} \langle \mathbf{Y}, \mathbf{\Sigma} \rangle \text{ s.t. } \text{tr}(\mathbf{Y}^t) = 1, \forall t \in [r],$$

$$\mathbf{Y} = \sum_{t \in [r]} \mathbf{Y}^t \preceq \mathbb{I}$$

$$Y_{i,j}^t = 0 \text{ if } Z_{i,t} = 0, \forall t \in [r], i, j \in [p],$$

$$\mathbf{Y} = \sum_{t=1}^r \mathbf{Y}^t, \text{Rank}(\mathbf{Y}^t) = 1, \forall t \in [r].$$

PROPOSITION 1. Consider  $r$  matrices,  $\mathbf{Y}^t \in \mathcal{S}_+^p$ , such that  $\text{tr}(\mathbf{Y}^t) = 1$  and  $\text{Rank}(\mathbf{Y}^t) = 1$ . Then,  $\sum_{t \in [r]} \mathbf{Y}^t \preceq \mathbb{I}$  if and only if  $\langle \mathbf{Y}^t, \mathbf{Y}^{t'} \rangle = 0 \forall t, t' \in [r]: t \neq t'$ .

# Sparse and low-rank reformulation

$$\begin{aligned} \max_{\substack{\mathbf{Z} \in \{0,1\}^{p \times r}: \mathbf{Y} \in \mathcal{S}^p, \mathbf{Y}^t \in \mathcal{S}_+^p \\ \langle \mathbf{E}, \mathbf{Z} \rangle \leq k}} \max_{\mathbf{Y}} \langle \mathbf{Y}, \mathbf{\Sigma} \rangle \quad s.t. \quad & \mathbf{Y} \preceq \text{Diag} \left( \min \left( \mathbf{e}, \sum_t \mathbf{Z}_t \right) \right), \mathbf{Y} = \sum_{t=1}^r \mathbf{Y}^t, \\ & \text{tr}(\mathbf{Y}^t) = 1, \quad \forall t \in [r], \quad Y_{i,j}^t = 0 \text{ if } Z_{i,t} = 0 \quad \forall t \in [r], i, j \in [p], \\ & \text{Rank}(\mathbf{Y}^t) = 1, \quad \forall t \in [r]. \end{aligned}$$

## Next steps (in the paper)

- Relaxation: SDP/SOC for rank, big-M for logical constraints
- Strengthened formulation via symmetry-breaking inequalities
- Strengthened formulation via valid inequalities linking  $\mathbf{Y}^t$  and  $\mathbf{Z}_t$
- Relax-round-then-estimate heuristics

# Valid Inequalities via Individual Sparsity Budgets

Suppose  $t^{th}$  component  $k_t$  sparse with  $\sum_t k_t =: k$ . Impose

$$\left( \sum_{j=1}^p |Y_{i,j}^t| \right)^2 \leq k_t Y_{i,i}^t Z_{i,t} \quad \forall i \in [p], \forall t \in [r]$$
$$\sum_{i \in [p]: i \neq j} Y_{i,j}^t{}^2 \leq (k_t - 1) Z_{j,t} (Z_{j,t} - Y_{j,j}^t) \quad \forall j \in [p].$$

→ **Tightens** relaxation substantially but requires knowledge of  $k_t$ 's

# Feasible Solutions via Alternating Minimization

## A Lagrangean Relaxation of Sparse PCA

$$\max_{\mathbf{Z} \in \{0,1\}^{p \times r}: \langle \mathbf{e}, \mathbf{Z}_t \rangle \leq k_t} \max_{\mathbf{Y} \in \mathcal{S}_+^p, \mathbf{Y}^t \in \mathcal{S}_+^p} \sum_{t \in [r]} \langle \mathbf{Y}^t, \mathbf{\Sigma} \rangle - \lambda \sum_{t, t' \in [r]: t \neq t'} \langle \mathbf{Y}^t, \mathbf{Y}^{t'} \rangle \text{ s.t. } \text{tr}(\mathbf{Y}^t) = 1, \forall t \in [r],$$

$$Y_{i,j}^t = 0 \text{ if } Z_{i,t} = 0, \forall t \in [r], i, j \in [p],$$
$$\text{Rank}(\mathbf{Y}^t) = 1, \forall t \in [r].$$

$\lambda > 0$  a penalty parameter

Fix all but one PC  Single component sparse PCA

Alternating minimization  High-quality solutions

Code available on GitHub:  [ryancorywright/MultipleComponentsSoftware](https://github.com/ryancorywright/MultipleComponentsSoftware)

# Application I: pitprops dataset ( $p=13$ ), $r=6$ PCs

Lu and Zhang, MP (2011)

“we deduce that for the Pitprops data, it seems not possible to extract six highly sparse (e.g., around 60 zero loadings), nearly orthogonal and uncorrelated PCs while explaining most of variance as they may not exist.”

$r$	$k_t$	Alg. 1			Alg. 2			Branch-and-Bound			
		Obj.	Viol.	T(s)	Obj.	Viol.	T(s)	Obj.	Viol.	Nodes	T(s)
6	2	<b>0.749</b>	0	1.01	0.734	0	23.31	0.740	0.001	65600	> 600
	4	0.666	0	0.90	<b>0.814</b>	0.085	287.34	0.718	0.002	39200	> 600
	6	0.686	0	1.54	0.836	0.153	336.99	<b>0.723</b>	0.003	39600	> 600
	8	0.673	0	1.89	0.854	0.122	293.38	<b>0.745</b>	0.003	45700	> 600
	10	0.645	0	1.34	0.874	0.115	206.93	<b>0.790</b>	0.004	25200	> 600
ALSPCA-2			60			0.03			0.084		39.42
ALSPCA-3			63			0.00			0.222		65.97

Pitprops data

## Application II: Performance on UCI Datasets

- Fix  $k=15,30$ ,  $r=3$
- Compute worst bound, best AM solution over rank-3 allocations of sparsity budget

Dataset	$p$	$r$	$k$	Enumerated			
				$k_t$	Obj.	Rel. gap (%)	Viol.
Pitprops	13	3	15	(7, 6, 2)	0.595	3.67%	0.002
			30	(10, 10, 10)	0.650	0.28%	0.005
Ionosphere	34	3	15	(7, 6, 2)	0.299	0%	0
			30	(11, 10, 9)	0.402	2.35%	0
Geographical	68	3	15	(6, 5, 4)	0.221	0%	0
			30	(12, 12, 6)	0.420	0%	0
Communities	101	3	15	(6, 5, 4)	0.142	0.01%	0
			30	(11, 11, 8)	0.247	0.17%	0

The average semidefinite relaxation gap is less than 1%!

# Application III: US Senate Voting Patterns 2021-2022

## Problem setting

Explain voting patterns in 117<sup>th</sup> senate

- $n=100$  data points (senators)
- $p=839$  features (bills/amendments/noms...)
  - Screened out  $p=17$  unanimous votes
- Compute two 5-sparse PCs via SOC relax+greedy rounding, score senators

## Results

First PC

- Initial vote on Inflation Reduction Act (IRA)
- Another vote on IRA
- Lower cost of Insulin
- Another vote on IRA
- Amendment to IRA

The July 2022 "Inflation Reduction Act"  
All five votes on same day!

Second PC

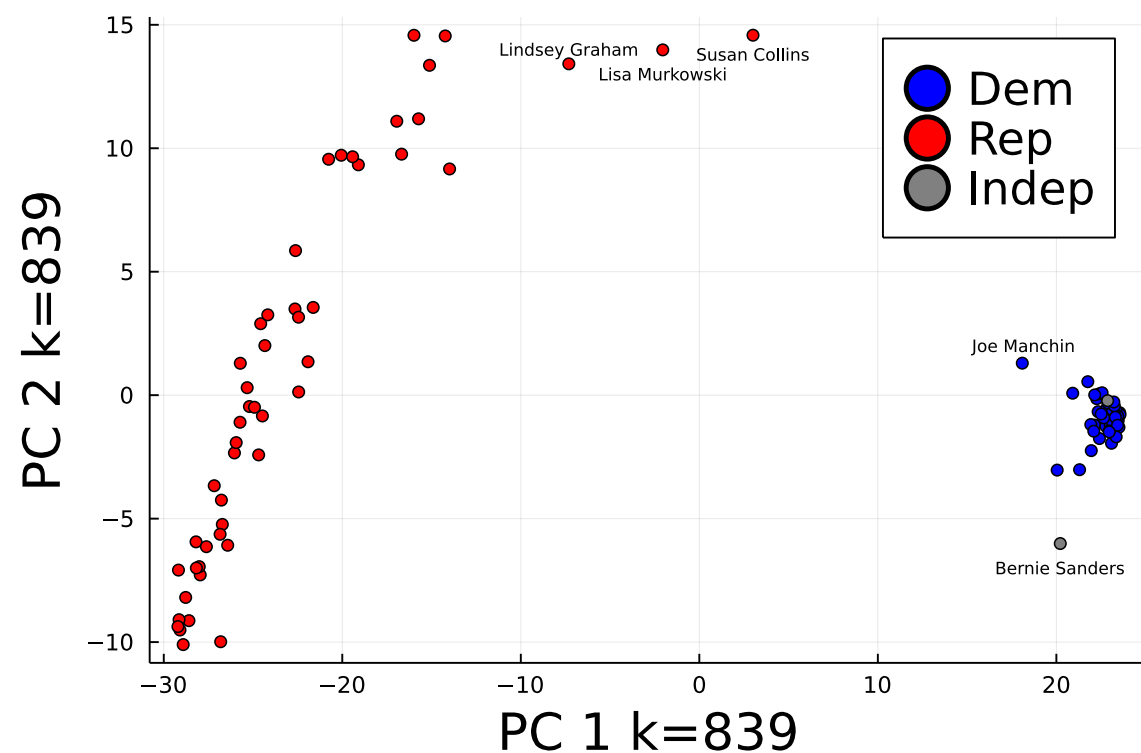
- Revenues from leasing oil/gas on federal land
- Extend Trump tax cuts
- Deficit neutral fund for catch-and-release
- Update text of 2021-22 budget
- Pass 2021-22 budget

The Feb 2021 "American Rescue Plan" and 3 proposed amendments  
All five votes on same day!

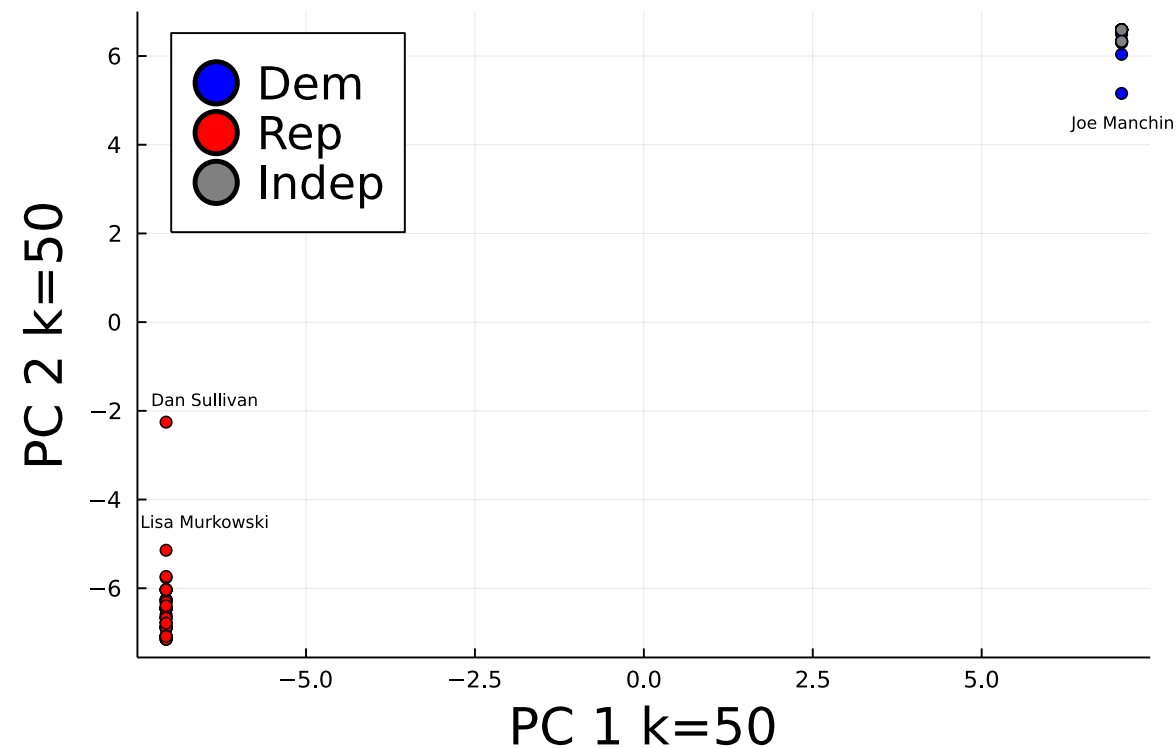
**Both PCs perfect classifiers by party!**  
**Plots not very interesting, so increase  $k$**

# Application III: US Senate Voting Patterns 2021-2022

Results no sparsity



Results k=50





# Conclusion

Sparse PCA with  $r > 1$  PCs remains a largely open problem

Sparse PCA with  $r > 1$  PCs is **significantly more challenging** than the  $r = 1$  case

Orthogonality  $\rightarrow$  rank constraints

## Semidefinite Relaxation + Alternating Minimization Solves Sparse PCA to (Near) Optimality

- Bound gaps of 1%-5% in practice, depending on application
- Key is to use good valid inequalities to tighten formulations

# Thank you!