Master Thesis

An ADMM approach to estimate Sparse Abnormal Subspace for Anomaly Detection and Interpretation
(基于ADMM的PCA异常空间稀疏化方法)

Supervisors:
Professor Wenguang CHEN
Professor Ying ZHAO

Kazi Abir Adnan(安奈威)
Student ID: 2014280162
Tsinghua University

GOALS

- ANOMALY DETECTION
- ANOMALY INTERPRETATION

OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
- Experiments
- Discussion
- Conclusion
- References

ANOMALY DETECTION (1/2)

- Most instances have normal trend
  - Very few fit least
  - Termed as anomaly
- Method to detect anomalies
  - Human annotations impossible
- Unlabeled dataset
  - Unsupervised method best choice
- PCA (Principal Component Analysis) can exploit the fact
  - Pioneering tool to find anomalies in high dimensions
  - Known as dimension reduction tool
  - Better visualization support (direction of eigenvectors)
  - Unsupervised
  - Focuses on normal trend of data explaining maximum variance

ANOMALY DETECTION (2/2)

- Downside of PCA
  - Have few zero entries on loadings
    - Non-sparse in nature
    - PCs cannot be interpreted
  - Interpreting PCA can be inherently challenging problem [RSR07]
  - Need sparse PC alongside dimension reduction

- Sparse PCA [ZHT06]
  - Make PCA understandable
  - Returns the most significant sparse PC covering maximum variance
  - Often we need more than one PC to explain required variance
  - No orthogonal guarantee

LITERATURE SHORTCOMINGS

- Cannot use existing methods directly in high dimensions
  - We focus on abnormal subspace
  - Abnormal subspace has more PCs
  - Introduces efficiency issues
  - Mostly focused on significant Principal Components
  - Capturing maximum variance
- Can be used only on Anomaly Detection
  - Anomaly Interpretation is infeasible
- Anomaly Interpretation is ignored on PCA-based model
  - Low sparsity
- Deflation technique
  - Increase in number of PCs increases computation time

OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
- Experiments
- Discussion
- Conclusion
- References

PRIME FOCUS

- Find a sparse Abnormal Subspace
  - Outliers have high projection
- Keep as simple as PCA
  - Keep variance equal
  - Keep PCs similar
  - Can go back to PCA
- Interpret anomalies
  - Using sparse abnormal PCs
- Reduce extravagant SDP computations
  - Introducing SDP formulation for subspace with convex relaxation
- Use efficient and effective optimization tool
  - Alternating Direction Method of Multipliers (ADMM)

ADMM (1/1)

- Popular tool/algorithm to solve convex optimization problem
- Solves optimization problems mostly in the form

$$\min_{x,z} f(x) + g(z)$$
$$s.t.\ Ax + Bz = c$$

  - f and g are convex
  - Two sets of variables with separate objective
- The augmented Lagrangian can be formed as,

$$L_\rho(x,z,y) = f(x) + g(z) + y^T(Ax + Bz - c) + \left(\frac{\rho}{2}\right)||Ax + Bz - c||_2^2$$

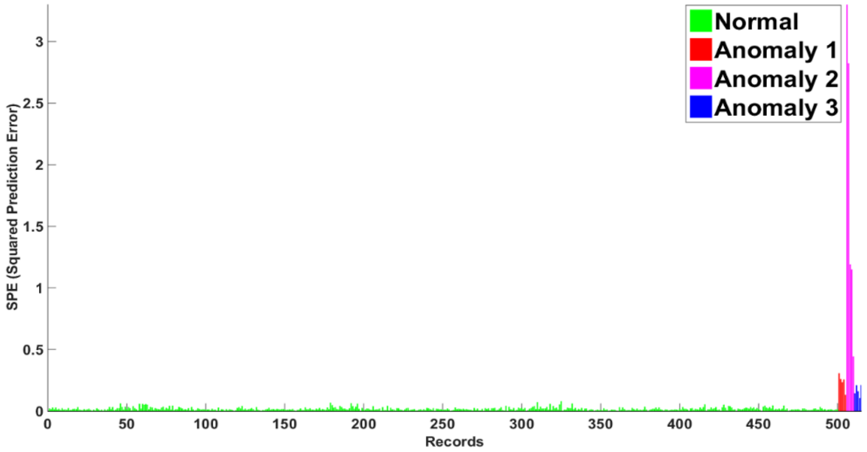
- ADMM Steps

$$x^{k+1} := argmin\, L_\rho(x, z^k, y^k) \quad \text{//x-minimization}$$
$$z^{k+1} := argmin\, L_\rho(x^{k+1}, z, y^k) \quad \text{//z- minimization}$$
$$y^{k+1} := y^k + \rho(x^{k+1} + B\, z^{k+1} - c) \quad \text{//Dual variable update}$$

**[BPC11]

ANOMALY DETECTION (1/1)

- $V_{abnormal} = (v_1, .., v_d)$ is *d*-dimensional abnormal subspace
  - Covers minimum variance
- $V_{normal} = (v_{d+1}, v_2, v_3, .., v_p)$ is the normal subspace
  - Defined by the significant *p* - *d* PCs
  - Complement of abnormal Subspace
  - Covers maximum variance
- For a data point y,
- its residual is
  - $\hat{y} = y - V_{normal} V_{normal}^T y$



- The squared length of projection residual length $\hat{y}$
  - Metric to indicate whether y is an anomaly
  - Squared Prediction Error (SPE)
  - The higher the SPE, it is more likely instance is an outlier

ANOMALY INTERPRETATION (1/1)

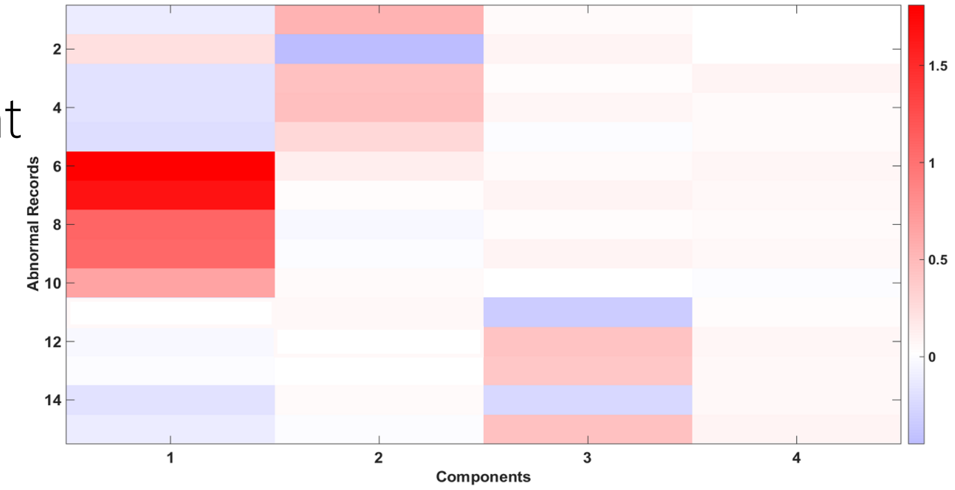
- Given V_{normal} and $V_{abnormal}$ subspace[BZS15]

$$\hat{y} = y - V_{normal}V_{normal}^{T}y = V_{abnormal}V_{abnormal}^{T}y$$

- Using orthogonal property [BZS15]

$$SPE = \hat{y}^{T}\hat{y} = \sum_{i=1}^{d}(v_i^{T}y)^2$$

- Squared sum of y's scalar projection on each abnormal PCs
- Abnormal PCs should be
  - Linear combination of important feature variables
  - Strong heatmap projection
  - Responsible for abnormality



RELATED WORK (1/2)

- FPS[VCLR13]
  - Emphasized on the span of eigenvectors called the 'Principal Subspace'
  - A novel convex optimization problem to estimate the d-dimensional principal subspace
  - Captures maximum variance representing normal data
  - Represents most significant sparse PC's
  - Introduces another issue of interpretation performance
- LFPCA[CL15]
  - Focused on sequential estimation of subspace
  - SDP with convex relaxation
  - Interpretation performance improves computing PCs sequentially
  - Enforced orthogonality constraint to be represented as eigenvectors

RELATED WORK (2/2)

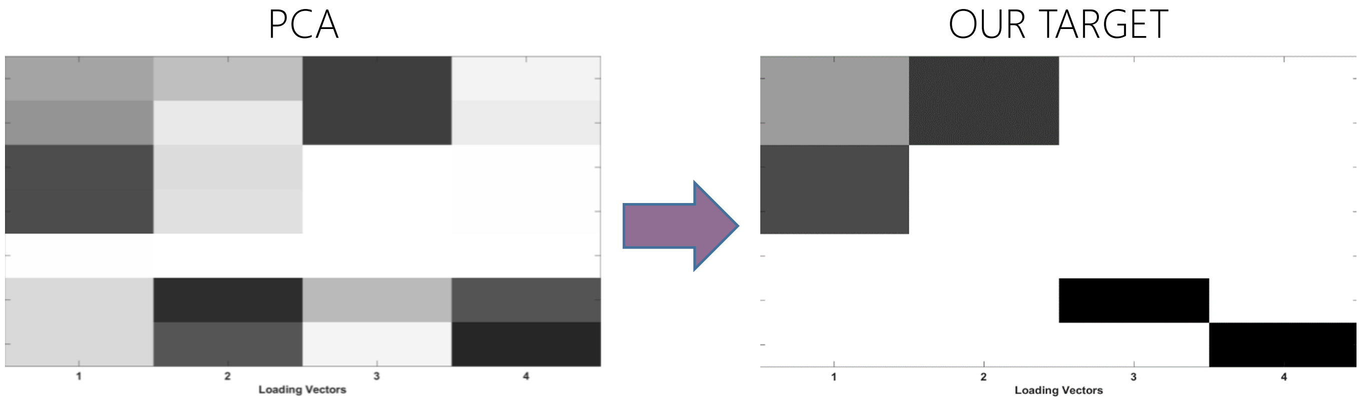
- ASPCA-BG[BZS15]
  - Estimates an 'Abnormal Subspace'
  - Represents anomalies with abnormal trend
  - Least significant PCs
  - Covers very minimum variance
  - Interprets anomalies
- Novel to all other mentioned related works in literature
  - All methods emphasizes on PCs with maximum variance
  - ASPCA-BG concerns about the PCs with lowest variance
- Normal data have almost zero length projection on abnormal subspace
- Uses Deflation technique
  - Solves SDP problem iteratively
  - Time consuming

COMPREHENSIVE COMPARISON

TASK	PCA	FPS[VCLR13]	LFPCA[CL15]	ASPCA-BG[BZS15]
ANOMALY DETECTION	✓	✓	✓	✓
SPARSITY	✗	✓	✓	✓
ANOMALY INTERPRETATION	✗	✗	✗	✓
LOW TIME CONSUMPTION	✓	✓	✓	✗
ABNORMAL SUBSPACE ESTIMATION	✗	✗	✗	✓

CONTRIBUTION

- Estimate d-dimensional abnormal subspace simultaneously
  - Using the connection of PCA
- An ADMM solution of SDP problem with convex relaxation
  - Low rank matrix approximation
  - Sparsity constraint
  - Enforce orthogonality constraint on computed PCs
- Reduce extravagant computation time
- Abnormal Sparse loading vectors representing anomalies



OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
  - Simultaneous Approach
  - Sequential Approach
- Experiments
- Discussion
- Conclusion
- References

SIMULTANEOUS ESTIMATION (1/4)

- Our focus is to find subspace explaining minimum variance
- The ordinary PCA problem with objective to capture the lowest variance

$$\min_{U\epsilon\, R^{p\times d}} Tr(U^T S U)$$
$$s.t. U^T U = I_d$$

- Quadratic problem to a linear problem
  - By considering $X = UU^T$, making X positive semidefinite

$$\min_{X,U} Tr(SX)$$
$$s.t. X = UU^T ; U^T U = I_d$$

- Or equivalently

$$\min_{X} Tr(SX)$$
$$s.t. X \text{ is a } d - \text{dimensional projection matrix}$$

SIMULTANEOUS ESTIMATION (2/4)

- Good News
  - Estimates Subspace
  - Simultaneous extraction of PCs
- Bad News
  - Finding the global minimizer is computationally demanding
- Solution
  - Formulate it with convex relaxation

SIMULTANEOUS ESTIMATION (3/4)

- $Tr(SX)$ is linear in X
  - Equivalent to a convex problem

$$\min_X \; Tr(SX)$$
$$s.t. X \in Convhull(all\ d - dim\ projection\ matrices)$$

- Add sparsity

$$\min_X Tr(SX) + \lambda\, ||X||_{1,1}$$
$$s.t. X \in Convhull(all\ d - dim\ projection\ matrices)$$

  - $||X||_{1,1}$ – Entrywise l_1 norm
  - λ – tuning parameter
- The tightest convex relaxation can be Fantope [VCLR13]

$$F_{p_d}(X) := \{ X : X = X^T, 0 \preccurlyeq X \preccurlyeq I \ and\ Tr(X) = d \}$$

  - Is a convex body [Fillmore & Williams 71]
  - Comprises with extreme points of convex hulls
  - Has non-zero eigenvalues
  - Orthonormal matrices with particular dimension

SIMULTANEOUS ESTIMATION (4/4)

- Our Abnormal Subspace estimation objective function

$$\min_{X} Tr(SX) + \lambda \, ||X||_{1,1}$$

$$s.t. X \epsilon \, F_{p_d}$$

- Solve a linear function over intersection of two convex sets
  - The l_1 ball
  - The Fantope, F_{p_d}
- How to exploit the sparsity constraint and low rank approximation constraint together?
  - Use ADMM
  - We propose ASPCA-ADMM algorithm

ASPCA-ADMM (1/4)

- Equivalent ADMM formulation
  - Introducing convex Indicator function $1_{F_{p_d}}(X)$

$$\min_X Tr(SX) + 1_{F_{p_d}}(X) + \lambda \,||X||_{1,1}$$

$$s.t. X \epsilon\, F_{p_d}$$

  - Splitting to two variables

$$\min_X \infty . 1_{F_{p_d}}(X) + Tr(SX) + \lambda \left||Y|\right|_{1,1}$$

$$s.t. X - Y = 0$$

- The augmented Lagrangian associated with it has the form

$$L\rho(X, Y, U) = \infty \cdot 1_{F_{p_d}}(X) + Tr(SX) + \lambda||Y||_{1,1} + \frac{\rho}{2}\left(||X - Y - U||_2^2 + ||U||_2^2\right)$$

  - Where $U = (1/\rho)Z$ is the scaled ADMM dual variable[BPC11]
  - ρ is the ADMM penalty parameter

ASPCA-ADMM (2/4)

- Iteration scheme is simple

$$X^{new} \leftarrow P_{F^d}(Y^{old} - U^{old} - {}^{S}/_{\rho})$$ ◃ Fantope projection

$$Y^{new} \leftarrow S_{\lambda/\rho}(X^{new} + U^{old})$$ ◃ Soft-thresholding

$$U^{new} \leftarrow U^{old} + \rho(X^{new} - Y^{new})$$ ◃ Dual update

- Then
  - $(X^{old}, Y^{old}, U^{old}) \rightarrow (X^{new}, Y^{new}, U^{new})$
  - Repeat until convergence is observed
- Fantope projection ≈ Soft thresholding singular values
- Use common patterns to solve minimization steps
  - Proximal Algorithms [PB14]
- Some algebraic manipulation reduces X and Y updates to
  - Computing Proximal Operators
    - X- update to Euclidean projection onto Fantope
    - Y- update to elementwise soft thresholding

ASPCA-ADMM (3/4)

- Fantope Projection can be computed
  - Euclidean projection onto $\boldsymbol{Fantope}$ and a closed form [VCLR13]

$$P_{F^d}(X) = \sum_i \gamma_i^+(\theta) u_i u_i^T$$

  - Where, $\gamma_i^+(\theta) = min(max(\gamma_i - \theta, 0), 1)$ and θ satisfies the equation $\sum_i \gamma_i^+(\theta) = d$.
- The second step is called the shrinkage step
  - Defined as Soft Thresholding for any λ/ρ

$$S_{\lambda/\rho}(X) = sign(X)\max(|X| - \lambda/\rho, 0)$$

- And we chose to terminate our ADMM when

$$max(|||X^{(t)} - Y^{(t)}|||_2^2, \rho^2 ||| Y^{(t)} - Y^{(t-1)}|||_2^2) \leq \in$$

  - Keep the primal and dual residual norms within a constant factor

ASPCA-ADMM (4/4)

Algorithm 1 ASPCA-ADMM

Input: D, d , λ, ρ , ϵ
S ←covariance(D) ◃ Covariance matrix
$Y^{(0)} \leftarrow 0$, $U^{(0)} \leftarrow 0$ ◃ Initialization
repeat t = 0, 1, 2, 3, . . .
    $X^{(t+1)} \leftarrow P_{F^d}(Y^{(t)} - U^{(t)} - S/\rho)$ ◃ Fantope projection
    $Y^{(t+1)} \leftarrow S_{\lambda/\rho}(X^{(t+1)} + U^{(t)})$ ◃ Soft-thresholding
    $U^{(t+1)} \leftarrow U^{(t)} + X^{(t+1)} - Y^{(t+1)}$ ◃ Dual update
until $\max(|||X^{(t)} - Y^{(t)}|||_2^2, \rho^2|||Y^{(t)} - Y^{(t-1)}|||_2^2) \leq d\epsilon^2$ ◃Stopping criterion
return $Y^{(t)}$

OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
  - Simultaneous Approach
  - Sequential Approach
- Experiments
- Discussion
- Conclusion
- References

SEQUENTIAL APPROACH (1/2)

- Estimate Abnormal Subspace by computing the Abnormal PCs sequentially
  - One by one
- Individual eigenvectors have better interpretability [CL15]
  - Different eigenvectors may have different support
- Required/Guarantee
  - Sparsity
  - Orthogonality

SEQUENTIAL APPROACH (2/2)

- Two major steps
- First step
  - The first least significant eigenvector is simple
    - $\hat{X}_1 = \min_{X_1 \in P_{F^1}} Tr(SX_1) + \lambda \, ||X_1||_{1,1}$
    - $\hat{v}_1 = first\ leading\ eigenvector\ of \hat{X}_1$
- Second step
  - For $j \ \geq 2$, let $\hat{V}_{j-1} = (\hat{v}_1, \ldots, \hat{v}_{j-1})$
    - Deflated Fantope : $DF_{p_j} \ = \{ X_j \in F_{P_1};\ X_j \hat{V}_{j-1} \ = \ 0\}$
    - $\hat{X}_j = \min_{X_j \in DF_{p_j}} Tr(SX_j) + \lambda \, ||X_j||_{1,1}$
    - $\hat{v}_j = first\ leading\ eigenvector\ of \hat{X}_j$
- Guarantees orthogonality of estimated eigenvectors
- ASPCA-ADMMSEQ

ASPCA-ADMMSEQ (1/3)

- ADMM iterations
  - Almost same as previous approach
- Fantope Projection can be computed setting $\boldsymbol{d}$ =1 [VCLR13]

$$P_{F^1}(X) = \sum_i \gamma_i^+(\theta) u_i u_i^T$$

  - Where, $\gamma_i^+(\theta) = min(max(\gamma_i - \theta, 0), 1)$ and θ satisfies the equation $\sum_i \gamma_i^+(\theta) = 1$.
- Deflated Fantope projection can be computed as [CL15]
  - Frobenius norm projection onto Deflated Fantope
  - Let $\widehat{U}_{j-1} \in R^{p\times p-j+1}$be an orthonormal complement basis of $\widehat{V}_{i-1}$, then

$$P_{D_j}(X) = \widehat{U}_{j-1}\left[\sum_{i=1}^{p-j+1} \gamma_i^+(\theta)\eta_i\eta_i^T\right]\widehat{U}_{j-1}^T$$

    - Where, (γ_i, η_i) are the Eigen components of $\widehat{U}_{j-1}^T X \widehat{U}_{j-1}$
    - $\gamma_i^+(\theta) = min(max(\gamma_i - \theta, 0), 1)$,
    - And θ is chosen such that $\sum_{i=1}^{p-j+1} \gamma_i^+(\theta) = 1$
- The second step is same as previous approach, Soft-thresholding

$$S_{\lambda/\rho}(X) = sign\,(X)\ \max(|X| - \lambda/\rho\ , 0)$$

ASPCA-ADMMSEQ (2/3)

- 2 stopping criterion for 2 loops
- An Outer loop controls required number of Abnormal PCs
  - Terminate sequential computation while $\boldsymbol{d}$ achieved
- An Inner loop controls ADMM convergence for each PC
  - And we chose to terminate our ADMM like previous approach
    - Keep the primal and dual residual norms within a constant factor,$\boldsymbol{\epsilon}$

ASPCA-ADMMSEQ (3/3)

Algorithm 2 ASPCA-ADMMSEQ

Input: D, d, λ, ρ, ϵ , j = d

S = covariance (D) ◃ Compute covariance matrix

$P = 0$ ◃ Rank-d Projection Matrix

$Y_1^{(0)} \leftarrow 0, U_1^{(0)} \leftarrow 0$ ◃ Initialization

repeat $t = 0, 1, 2, \ldots$. ◃ First Eigen Vector

$X_1^{(t+1)} \leftarrow P_{F_1}(Y_1^{(t)} - U_1^{(t)} - {}^{S}/_{\rho})$ ◃ Rank-1 Fantope projection

$Y_1^{(t+1)} \leftarrow S_{\lambda/\rho}(X_1^{(t+1)} + U_1^{(t)})$ ◃ Soft-thresholding

$U_1^{(t+1)} \leftarrow U_1^{(t)} + X_1^{(t+1)} - Y_1^{(t+1)}$ ◃ Dual update

until $\max(|||X_1^{(t)} - Y_1^{(t)}|||_2^2, \rho^2 |||Y_1^{(t)} - Y_1^{(t-1)}|||_2^2) \leq d\epsilon^2$ ◃Stopping criterion

$P = P \cup Y_1^{(t)}$ ◃ Add the first eigenvector

for $j = 2, \ldots, d$ ◃ compute ***d***-1 eigenvectors

$Y_j^{(0)} \leftarrow 0, U_j^{(0)} \leftarrow 0$ ◃ Initialization

repeat $t = 0, 1, 2, \ldots$

$X_j^{(t+1)} \leftarrow P_{DF_j}(Y_j^{(t)} - U_j^{(t)} - {}^{S}/_{\rho})$ ◃ Deflated Fantope projection

$Y_j^{(t+1)} \leftarrow S_{\lambda/\rho}(X_j^{(t+1)} + U_j^{(t)})$ ◃ Soft-thresholding

$U_j^{(t+1)} \leftarrow U_j^{(t)} + X_j^{(t+1)} - Y_j^{(t+1)}$ ◃ Dual update

until $\max(|||X_j^{(t)} - Y_j^{(t)}|||_2^2, \rho^2 |||Y_j^{(t)} - Y_j^{(t-1)}|||_2^2) \leq d\epsilon^2$ ◃Stopping criterion

$P = P \cup Y_j^{(t)}$ ◃ Add the j^{th} eigenvector

end

return P

OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
- Experiments
- Discussion
- Conclusion
- References

EXPERIMENTS

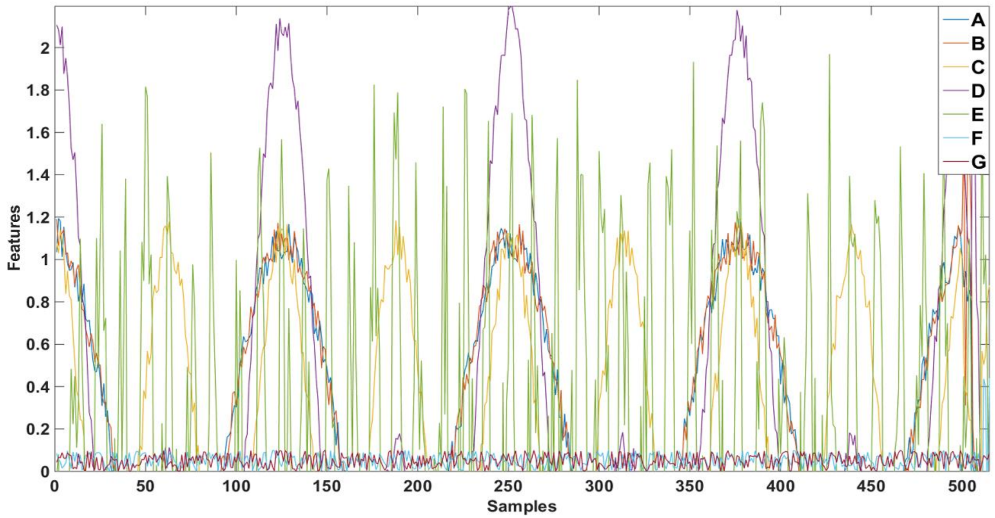
- Explain detection and interpretation process with a Synthetic dataset [BZS15]
  - 500 normal records and 15 anomalies
  - 3 types of anomalies
- Demonstrate performance on Real World benchmark dataset
  - Breast Cancer Dataset
    - 357 benign records (normal), 10 malignant records(abnormal)
  - KDD’99 Network Intrusion Dataset
    - 10% dataset
    - About 97242 normal records
    - 22 classes of attacks (abnormal)
    - Attacks can be classified into four main groups: DoS, Remote- to-local (R2L), User-to-root (U2R), and Probe

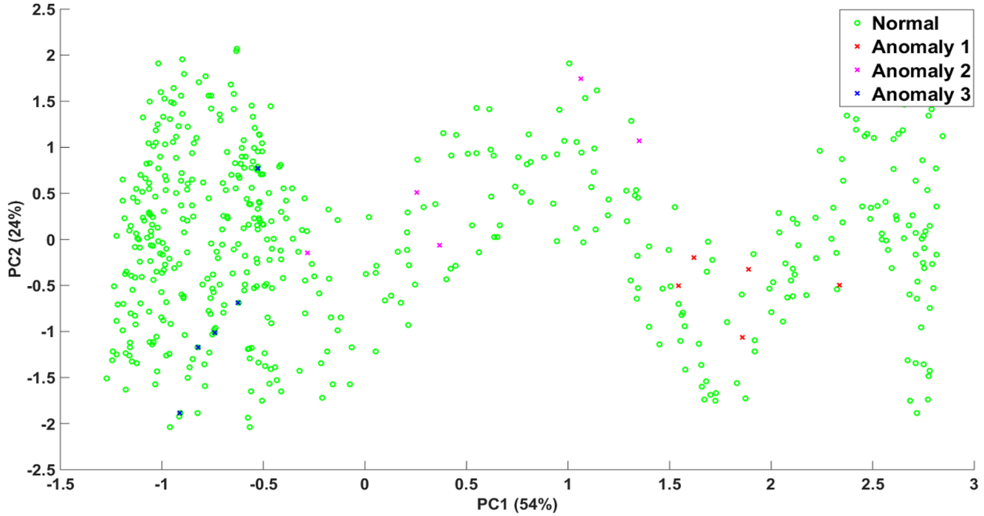
EVALUATION METIRCS

- Anomaly Detection
  - SPE
- Variance
  - Keep close to PCA
- Sparsity of loading vectors
  - Norm 1 ($||V||_{1,1}$)
  - $\text{Card}_{0.1}$
    - Number of entries with absolute value higher than 0.1
  - $\text{Card}_{0.01}$
    - Number of entries with absolute values higher than 0.01
- Anomaly Interpretation
  - Heatmap projection on few Abnormal PCs
  - Need human effort
  - Only ASPCA-ADMMSEQ's performance is illustrated

SYNTHETIC DATASET (1/7)

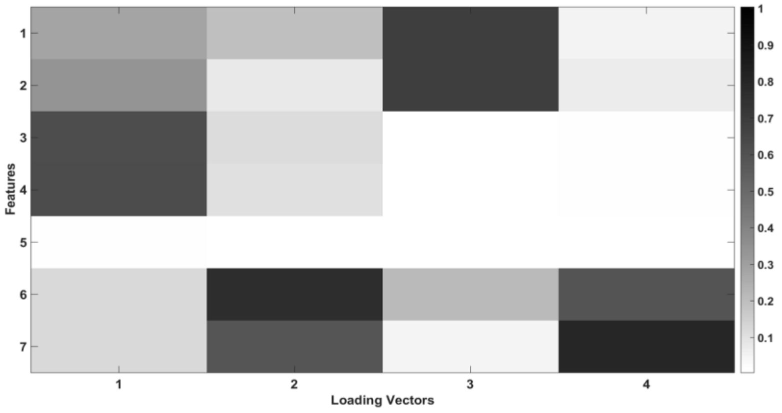
- Samples with Features
  - 500 normal records and 15 anomalies
  - 3 types of anomalies
- Feature Rules
  - D ≈ C + A
  - A ≈ B
  - F ≈ 0
  - G ≈ 0
- Anomalies break first three rules
- Projection on significant 2 PCs
- Variance captured ≈ 78%
  - PC1 (54%)
  - PC2 (24%)



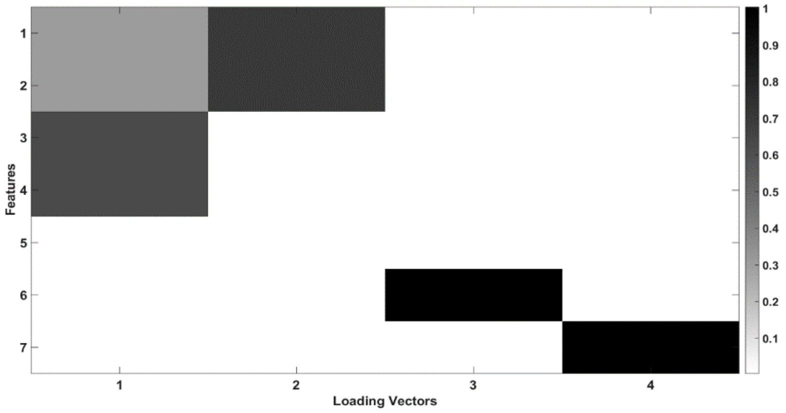


SYNTHETIC DATASET (2/7)

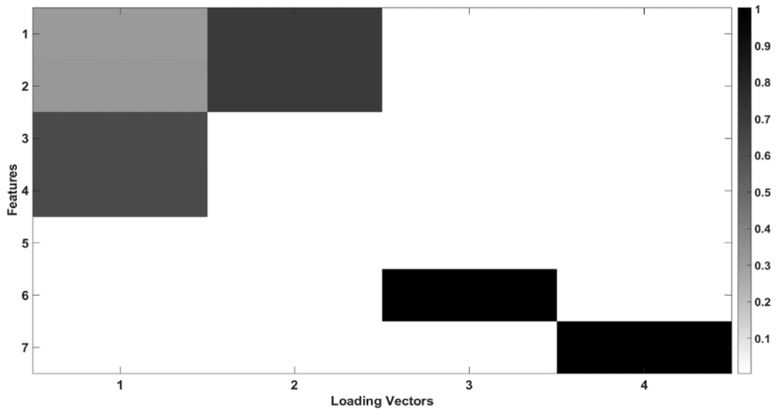
4 ABNORMAL PCs



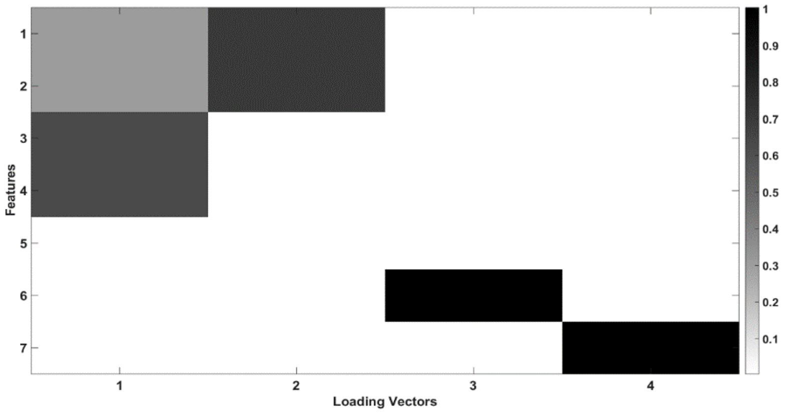
PCA



ASPCA-BG



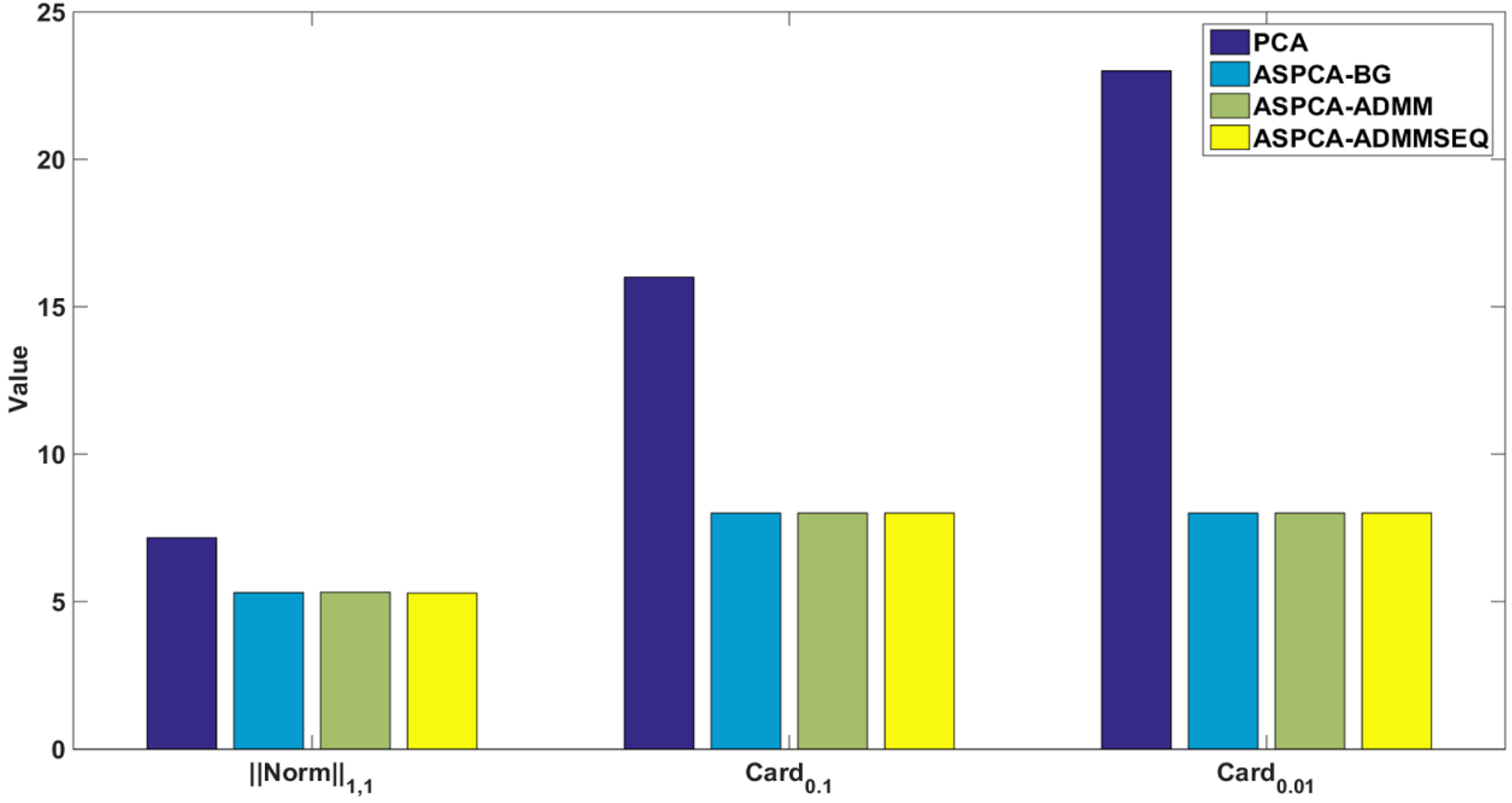
ASPCA-ADMM



ASPCA-ADMMSEQ

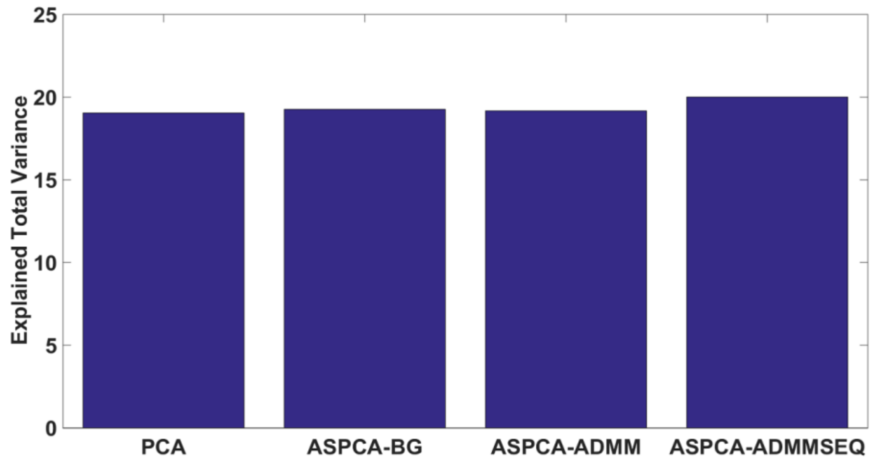
SYNTHETIC DATASET (3/7)

SPARSITY EVALUATION

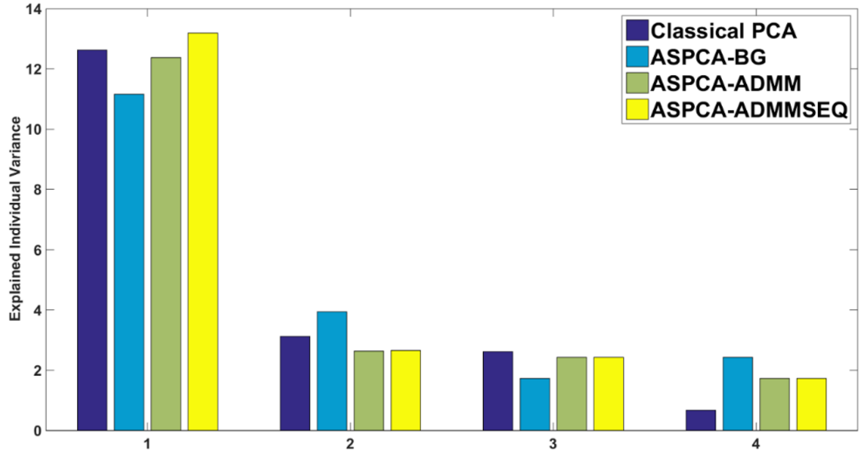


SYNTHETIC DATASET (4/7)

CAPTURED VARIANCE BY ABNORMAL SUBSPACE



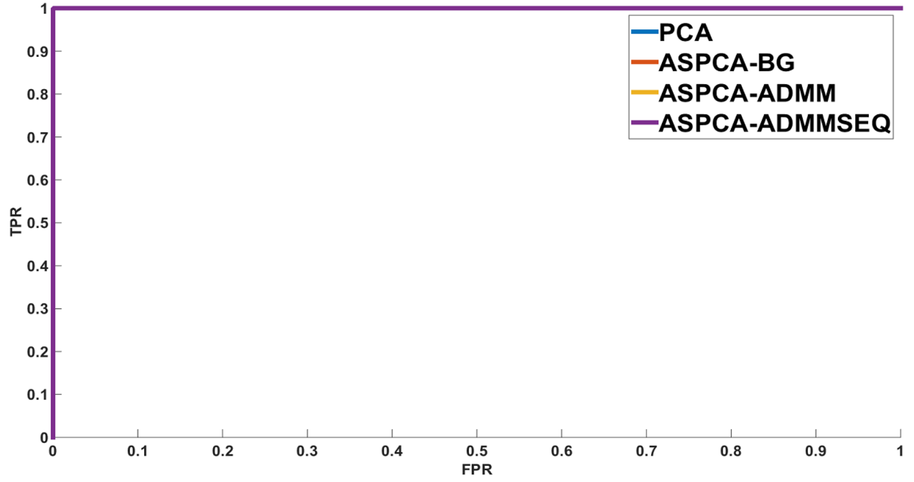
CAPTURED VARIANCE BY ABNORMAL PCs



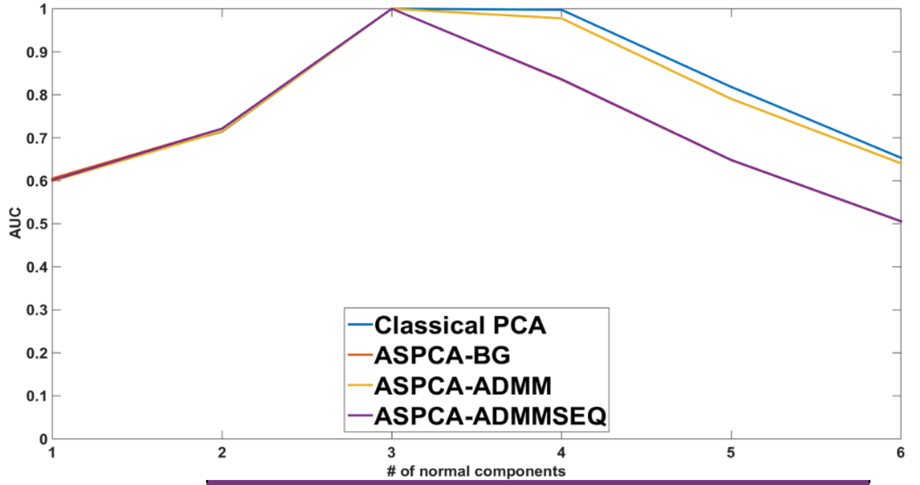
EXPERIMENTS (6/15)

SYNTHETIC DATASET (5/7)

PERFECT AUC FOR PCA, ASPCA-BG, ASPCA-ADMM,ASPCA-ADMMSEQ



SENSITIVITY OF #OF NORMAL COMPONENTS

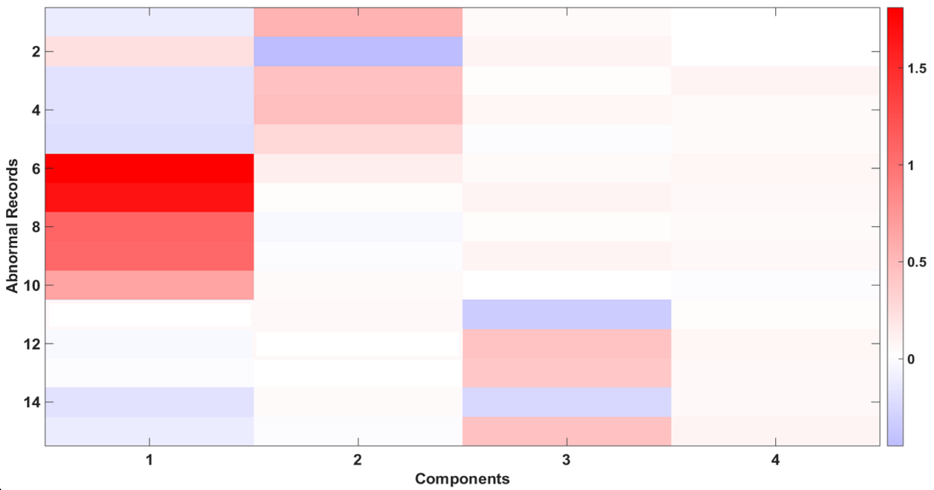


SYNTHETIC DATASET (6/7)

- Both ASPCA-ADMM & ASPCA-ADMMSEQ has same PCs
- Anomalies break rules
  - D ≈ C + A
  - A ≈ B
  - F ≈ 0
- Heatmap projection
  - Anomalies
  - Abnormal components

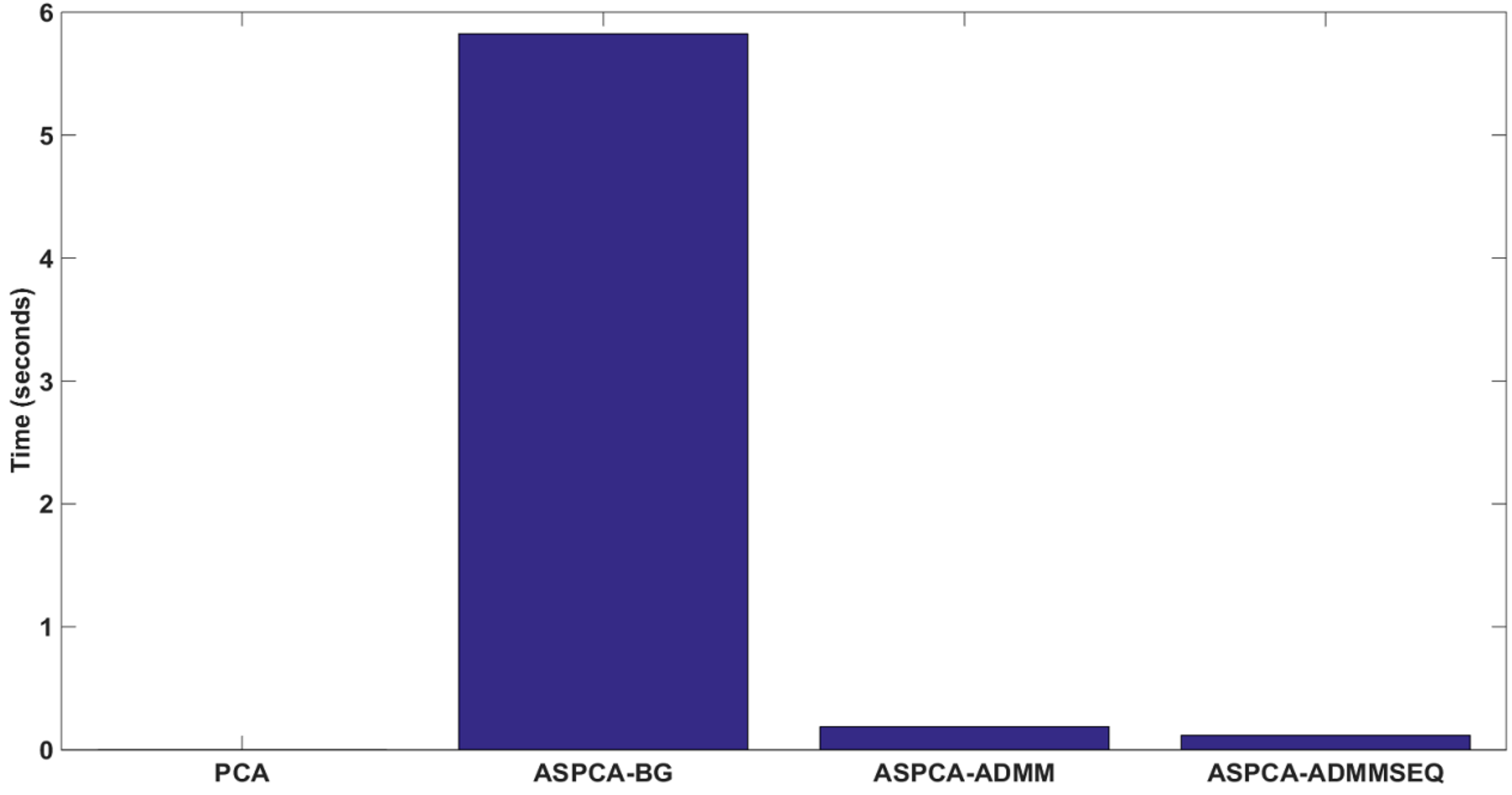
ABNORMAL PCs

Index	Components
1	-0.29148 A – 0.291663 B – 0.654865 C + 0.633345 D
2	-0.70433 A + 0.709883B
3	1 F
4	1 G



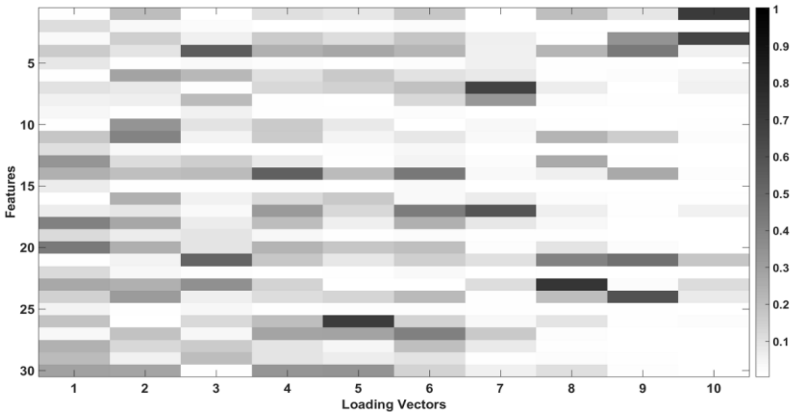
SYNTHETIC DATASET (7/7)

TIME CONSUMPTION

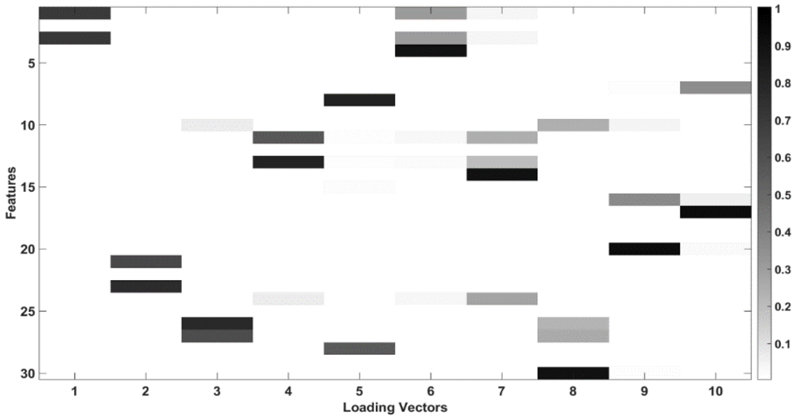


BREAST CANCER DATASET (1/4)

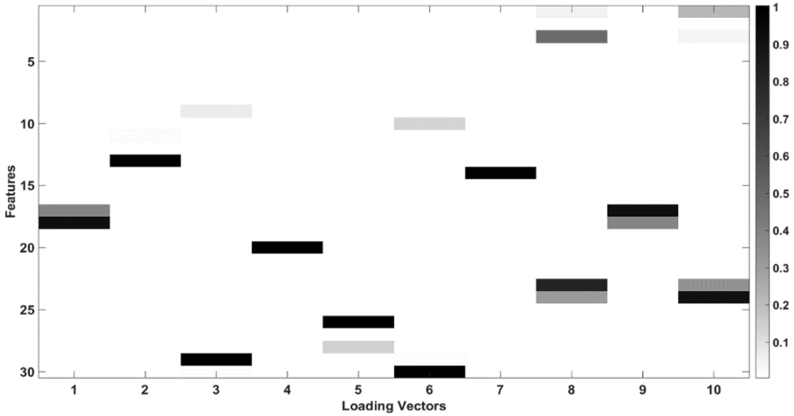
10 ABNORMAL PCs



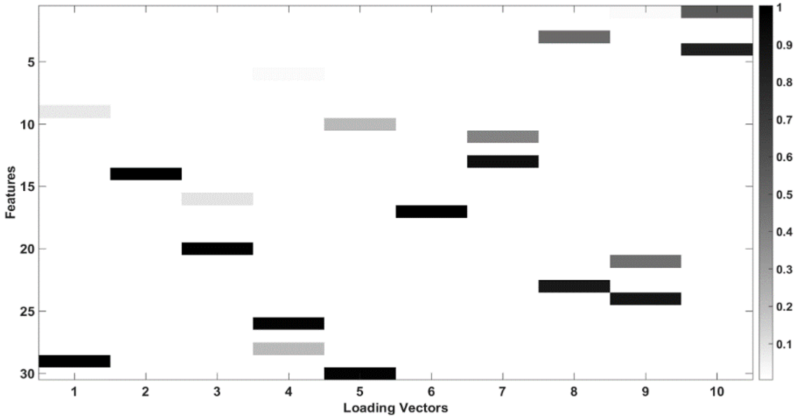
PCA



ASPCA-BG



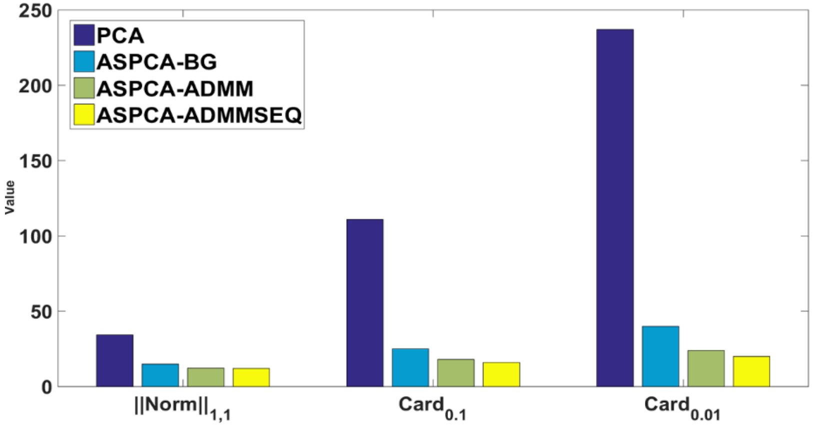
ASPCA-ADMM



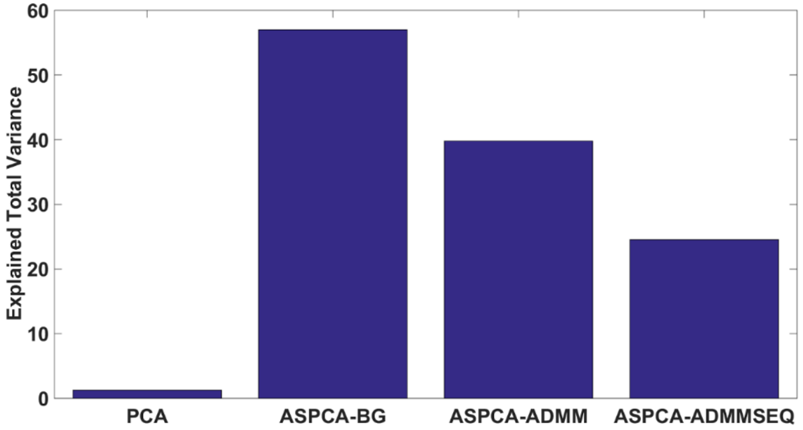
ASPCA-ADMMSEQ

BREAST CANCER DATASET (2/4)

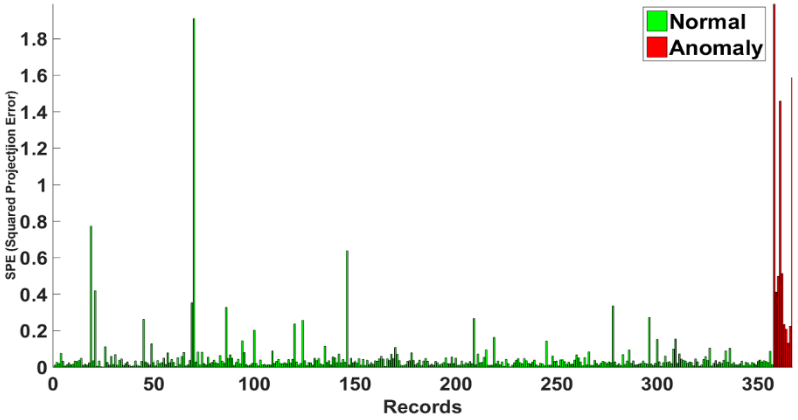
SPARSITY EVALUATION



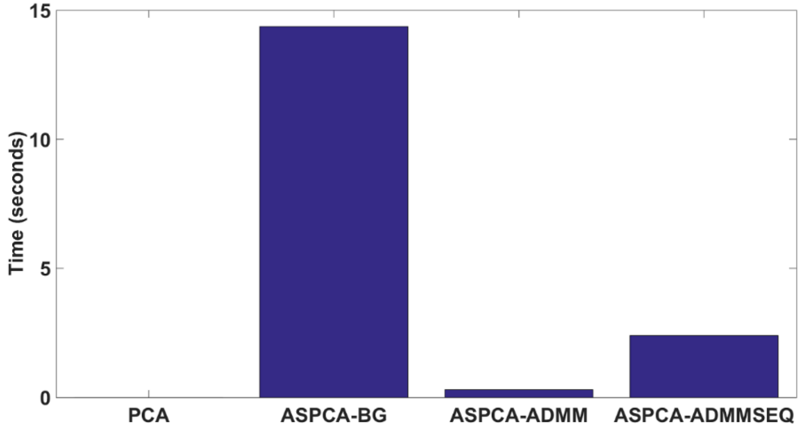
CAPTURED VARIANCE BY ABNORMAL SUBSPACE



SPE VALUES

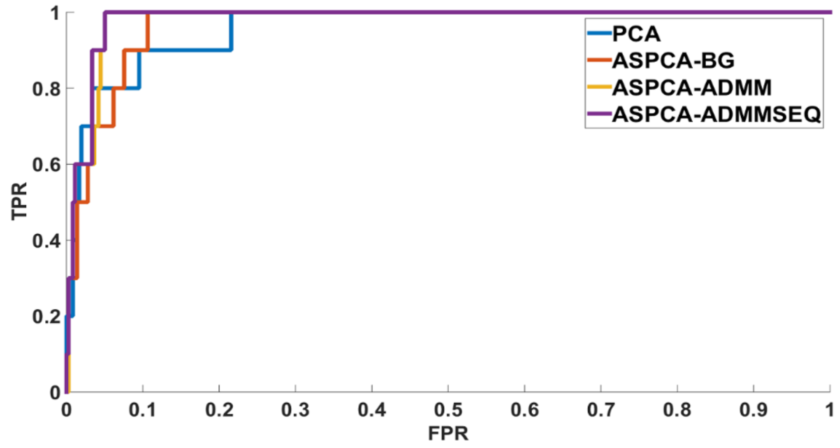


TIME CONSUMPTION

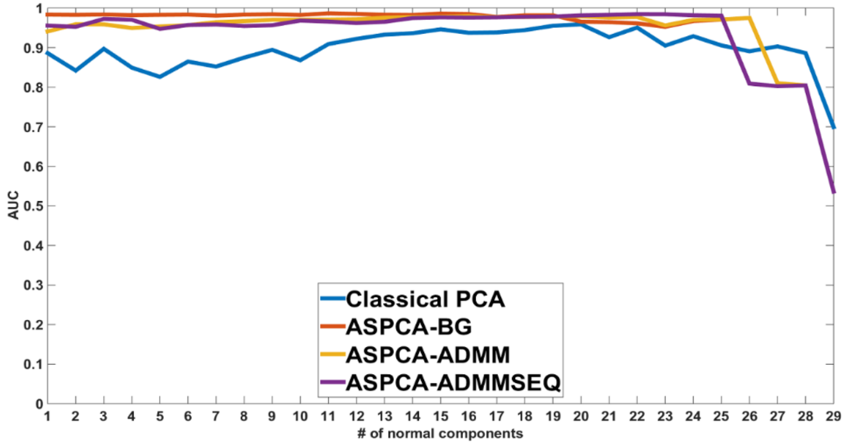


BREAST CANCER DATASET (3/4)

AUC COMPARISON



SENSITIVITY OF #OF NORMAL COMPONENTS



AUC VALUES

Method	AUC
PCA	0.958
ASPCA-BG	0.965
ASPCA-ADMM	0.978
ASPCA-ADMMSEQ	0.981

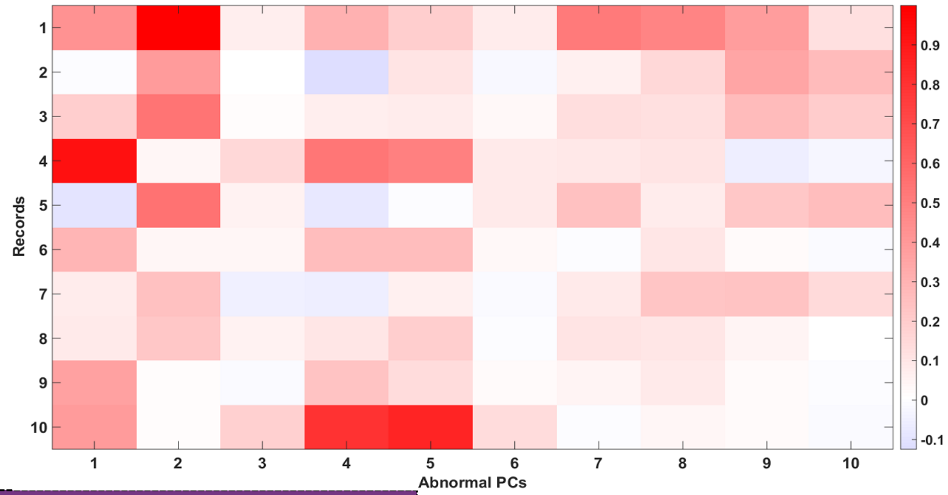
BREAST CANCER DATASET (4/4)

- 2 kinds of malignant records [BZS15] [WSH95]
  - area-se, area-worst, radius-worst
  - symmetry, fractal dimension, compactness

- Heatmap projection
  - Anomalies
  - Abnormal components

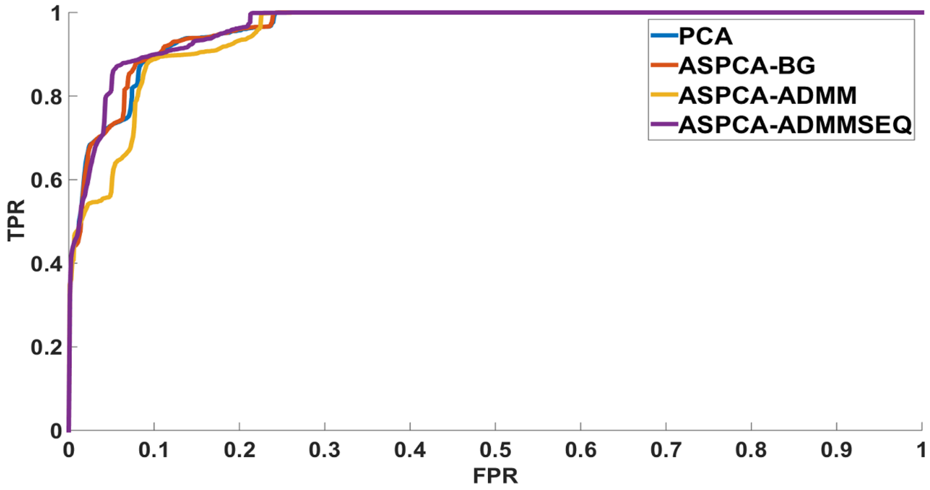
ABNORMAL PCs

Index	Components
1	- 0.08symmetry-mean + 0.996symmetry-worst
2	1area-se
4	- 0.02 compactness-mean + 0.97 compactness-worst - 0.20 concave-points-worst
5	- 0.21 fractal dimension-mean + 0.97 fractal dimension-worst
7	- 0.39 radius-se + 0.91 perimeter-se
8	- 0.49 perimeter-mean + 0.86 perimeter-worst
9	- 0.02 radius-mean - 0.47 radius-worst + 0.87 area-worst

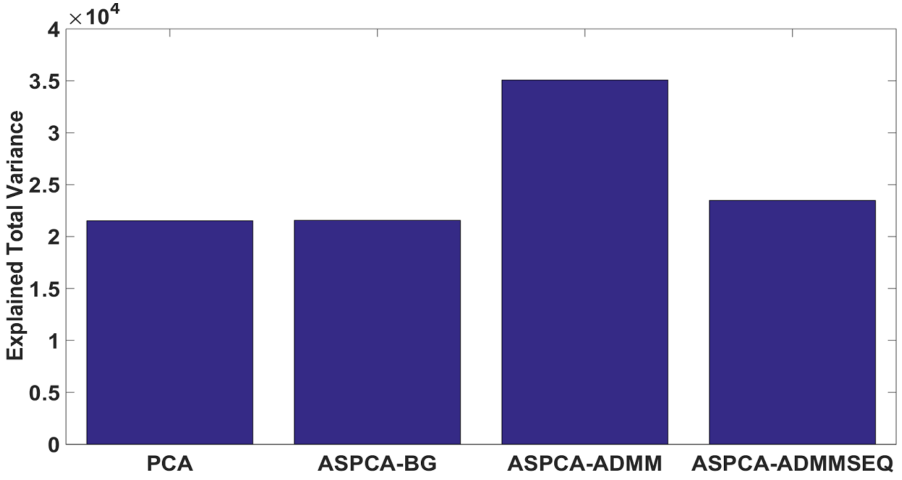


KDD'99 DATASET (1/1)

DETECTION EVALUATION



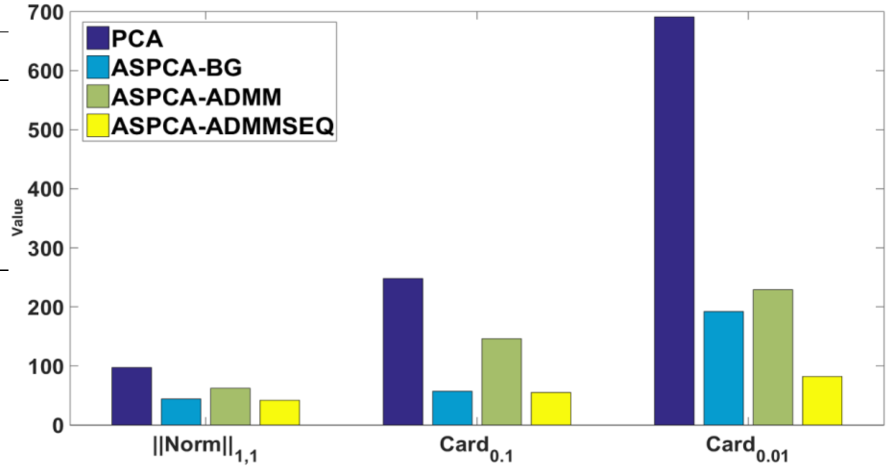
CAPTURED VARIANCE BY ABNORMAL SUBSPACE



AUC VALUES

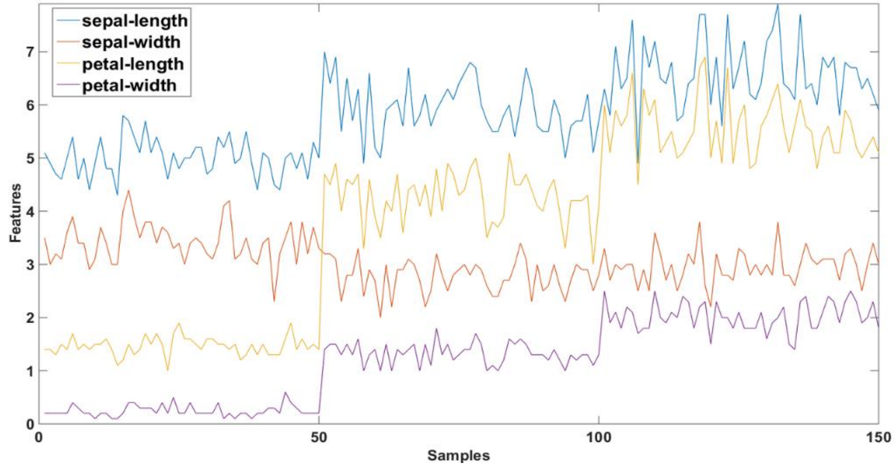
Method	AUC
PCA	0.962
ASPCA-BG	0.963
ASPCA-ADMM	0.952
ASPCA-ADMMSEQ	0.967

SPARSITY EVALUATION

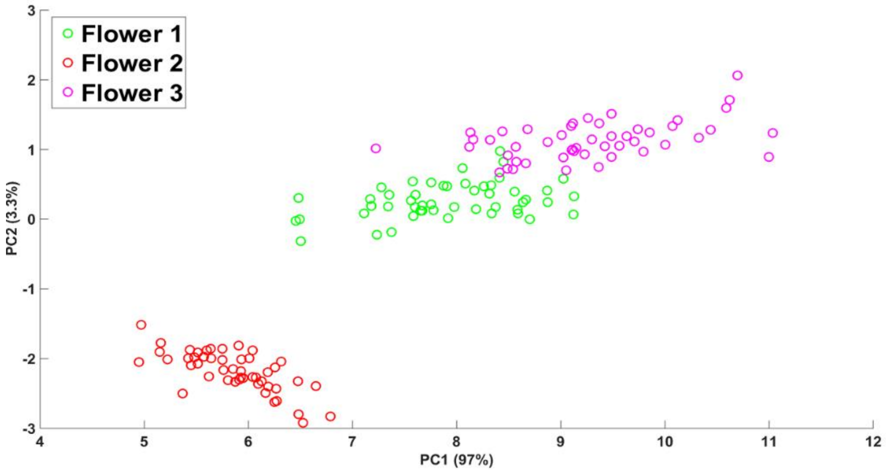


CLUSTERING PERFORMANCE?

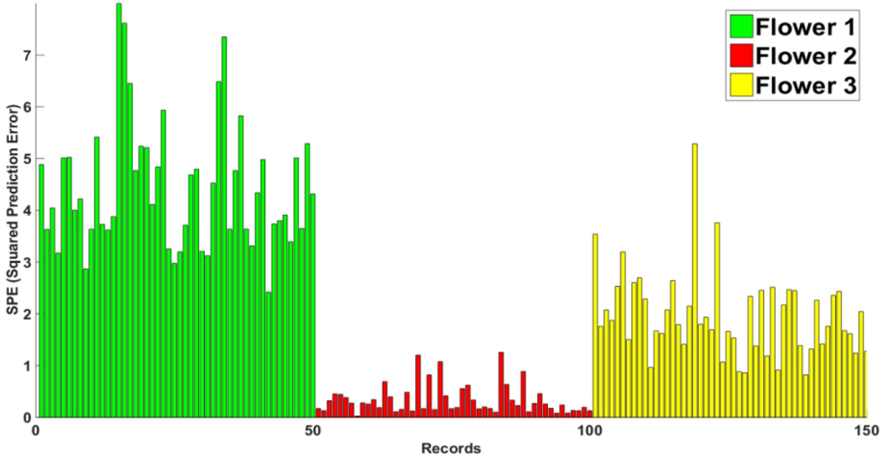
SAMPLES WITH FEATURES



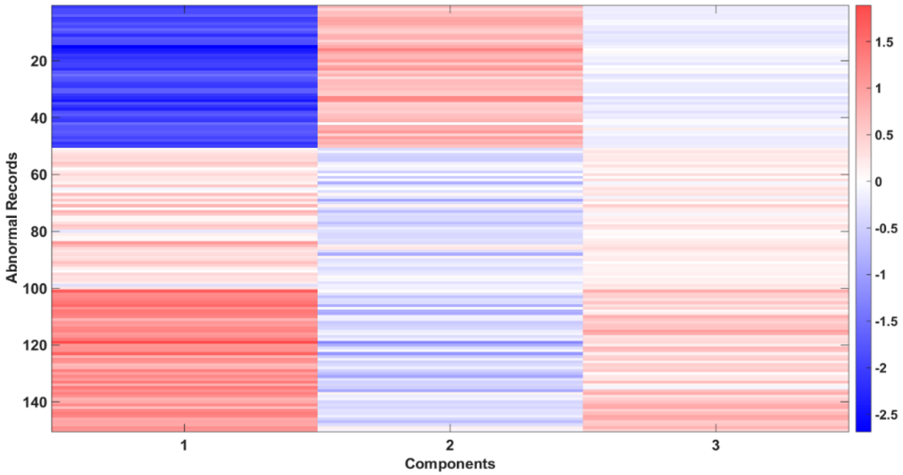
PROJECTION ON SIGNIFICANT 2 PCS



SPE VALUES ON ABNORMAL SUBSPACE



HEATMAP PROJECTION ON ABNORMAL PCs



OUTLINE

DISCUSSION (1/1)

- Tradeoff between sparsity, variance, AUC
- Sequential estimation of PCs
  - Has better interpretation performance
  - Keeps most of the correlations of variables
  - Captures variance like PCA
  - Formulate PCA result on certain circumstance
- Simultaneous estimation of PCs
  - Faster
  - Sparsity on subspace rather individual PCs
  - Adds additional variance to make sparse
  - Can not go back to PCA

OUTLINE

- Introduction
- Motivation
- Abnormal Subspace Estimation
- Experiments
- Discussion
- Conclusion
- References

SUMMARY (1/1)

- Proposed two models
  - ASPCA-ADMM
    - Simultaneous computation of eigenvectors
  - ASPCA-ADMMSEQ
    - Sequential computation of eigenvectors
- Computation time reduced
- Better Detection performance
  - Individual anomalies
- Better Sparsity performance
  - Understandable Anomaly Interpretation
- ASPCA-ADMMSEQ throws promising competency

FUTURE WORKS (1/1)

- Make Abnormal Subspace size, d non-parametric
- Tune other parameters automatically
  - penalty parameter ρ and regularization parameter λ
- Improve interpretation
  - Make it sparser on high dimensional datasets
- Projection on other convex bodies
  - We only used Fantope
- Make the model more robust and scalable for different big data problems
- Proper understanding of the behavior of data
- Clustering performance on other datasets

OUTLINE

REFERENCES I

- X.Bin, Y. Zhao, and B. Shen.
  - Abnormal Subspace Sparse PCA for Anomaly Detection and Interpretation.
  - ACM SIGKDD Workshop on ODDx3: Outlier Definition, Detection, and Description, 2015.
- V. Q. Vu, J. Cho, J. Lei, and K. Rohe
  - Fantope projection and selection: A near-optimal convex relaxation of sparse PCA.
  - In Advances in Neural Information Processing Systems, pp. 2670-2678, 2013.
- K. Chen, and J. Lei.
  - Localized Functional Principal Component Analysis.
  - Journal of the American Statistical Association, 110(511), pp.1266-1275, 2015.

REFERENCES II

- C. C. Aggarwal
  - Outlier analysis
  - In Data Mining, pp. 237-263. Springer International Publishing, 2015
- H. Zou, T. Hastie, and R. Tibshirani
  - Sparse Principal Component Analysis
  - Journal of Computational and Graphical Statistics, 15(2), pp.265-286, 2006
- Y. Zhang, A. d’Aspremont, and L. E. Ghaoui
  - Sparse PCA: Convex relaxations, algorithms and applications.
  - In Handbook on Semidefinite, Conic and Polynomial Optimization (pp. 915-940). Springer US, 2012.
- A. d'Aspremont, L. E. Ghaoui, Jordan, M.I. jordan, and G. R. G. Lanckriet.
  - A direct formulation for sparse PCA using semidefinite programming.
  - SIAM review, 49(3), pp.434-448, 2007.

REFERENCES III

- H. Ringberg, A. Soule, J. Rexford, and C. Diot
  - Sensitivity of PCA for traffic anomaly detection
  - In ACM SIGMETRICS Performance Evaluation Review (Vol. 35, No. 1, pp. 109-120). ACM, 2007
- W. H. Wolberg, W. Street, D. M. Heisey, and O. L. Mangasarian
  - Computer-derived nuclear features distinguish malignant from benign breast cytology
  - Human Pathology, 26(7), pp.792 -796, 1995
- S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein
  - Distributed optimization and statistical learning via the alternating direction method of multipliers.
  - Foundations and Trends® in Machine Learning, 3(1), pp.1-122, 2011.
- N. Parikh, and S.P. Boyd
  - Proximal Algorithms
  - Foundations and Trends in optimization, 1(3), pp.127-239, 2014.

Q&A

Thank You ☺