Heteroscedastic PCA (HPCA)

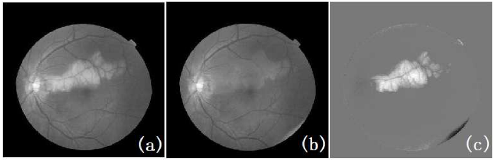
Javier Salazar Cavazos

Joint collaboration with Laura Balzano & Jeff Fessler

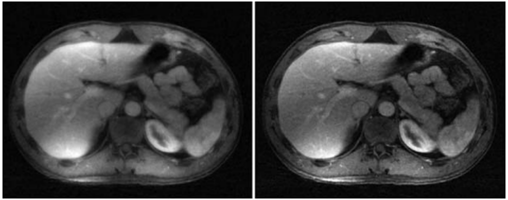
University of Michigan

January, 2023

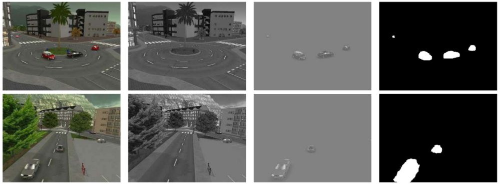
Applications



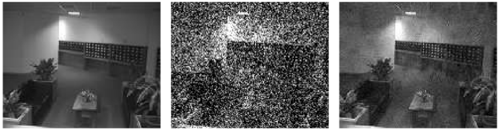
Lesion Detection [1]



Dynamic MRI Reconstruction [2]



Motion Estimation [3]

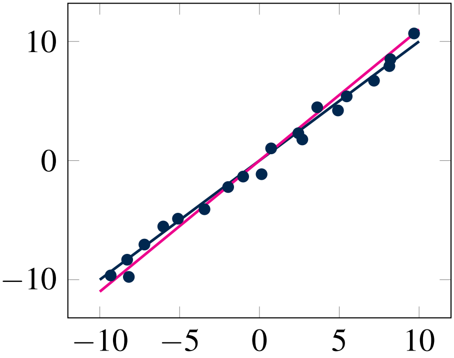


Image/Video Denoising [4]

Heteroscedasticity

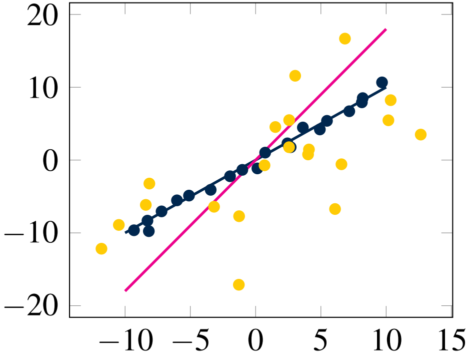
Homoscedastic Data

$\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon}$ s.t. $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \nu I)$



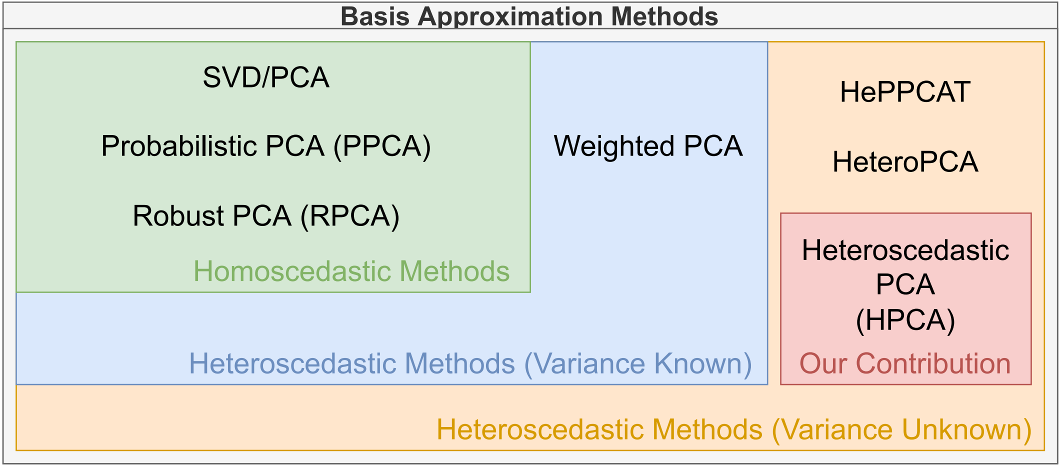
Heteroscedastic Data

$\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ s.t. $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \nu_i I)$



$\boldsymbol{x}_i = U\boldsymbol{z}_i$ where $\hat{U}$ = estimated subspace basis, $\boldsymbol{z}_i$ = basis coordinates

Background


Basis Approximation Methods
SVD/PCA
Probabilistic PCA (PPCA)
Robust PCA (RPCA)
Homoscedastic Methods
Weighted PCA
Heteroscedastic Methods (Variance Known)
HePPCAT
HeteroPCA
Heteroscedastic PCA (HPCA)
Our Contribution
Heteroscedastic Methods (Variance Unknown)

Robust PCA (RPCA)

Key Idea

Robust PCA = PCA + Outlier Robustness

Let $Y = [\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \,] \in \mathbb{R}^{D \times N}$ be a data matrix

Decompose Y such that $\underbrace{Y}_{\text{data matrix}} = \underbrace{L}_{\text{low rank data}} + \underbrace{S}_{\text{outlier matrix}}$

Solve the following optimization problem:

$$\arg\min_{L,S} \underbrace{\lambda_r \|L\|_*}_{\text{low rank}} + \underbrace{\|S\|_{1,1}}_{\text{outliers}} \quad \text{s.t. } Y = L + S \text{ where } \|L\|_* = \sum_{i=1}^{\min(M,N)} \sigma_i(L)$$

Model Limitations

$$\text{RPCA:} \quad \arg\min_{L,S} \underbrace{\lambda_r \|L\|_*}_{\text{low rank}} + \underbrace{\|S\|_{1,1}}_{\text{outliers}} \quad \text{s.t.} \quad Y = L + S \text{ where } \|L\|_* = \sum_{i=1}^{\min(M,N)} \sigma_i(L)$$

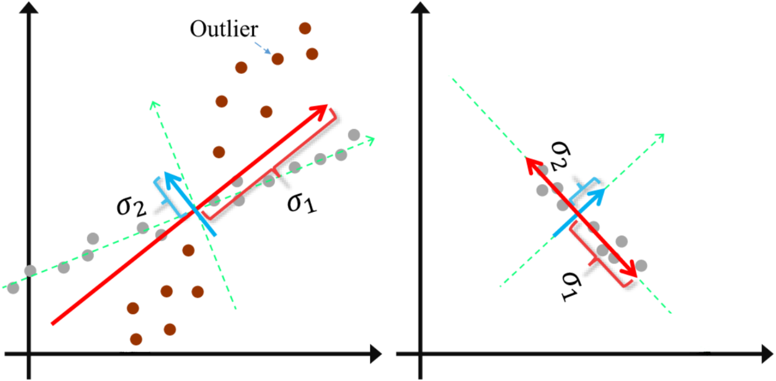


Image Courtesy: Oh et. al. [5]

Heteroscedastic PCA (HPCA)

Key Idea

HPCA = Robust PCA (RPCA) + Heteroscedastic Noise

Let $Y = [\, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_n \,] \in \mathbb{R}^{D \times N}$ and $\Pi = \text{diagm}(\nu_1, \ldots, \nu_n) \in \mathbb{R}^{N \times N}$

Decompose Y such that $\underbrace{Y}_{\text{data matrix}} = \underbrace{X}_{\text{low rank data}} + \underbrace{Z}_{\text{noise matrix}}$

Then the optimization problem we posed is the following:

$$\arg\min_{X,Z,\Pi} \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{\text{low rank}} + \underbrace{\frac{1}{2}\|Z\Pi^{-1/2}\|_F^2}_{\text{weighted noise}} + \underbrace{\frac{D}{2}\log \overbrace{|\Pi|}^{\text{det.}}}_{\text{unk. variance}} \quad \text{s.t. } Y = X + Z$$

Heteroscedastic PCA (HPCA)

Let $Y = [\, \boldsymbol{y}_1, ..., \boldsymbol{y}_n \,] \in \mathbb{R}^{D \times N}$ and $\Pi = \text{diagm}(\nu_1, \ldots, \nu_n) \in \mathbb{R}^{N \times N}$

$$\underset{X,Z,\Pi}{\arg\min} \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{\text{low rank}} + \underbrace{\frac{1}{2}\|Z\Pi^{-1/2}\|_F^2}_{\text{weighted noise}} + \underbrace{\frac{D}{2}\log \overbrace{|\Pi|}^{\text{det.}}}_{\text{unk. variance}} \quad \text{s.t. } Y = X + Z$$

$$\|X\|_{\circledast,\alpha} \triangleq \sum_{i=\alpha+1}^{\min(M,N)} \sigma_i(X) = \sum_{i=1}^{\min(M,N)} \sigma_i(X) - \sum_{i=1}^{\alpha} \sigma_i(X) = \|X\|_* - \|X\|_{\text{Ky-Fan}(\alpha)}$$

Note!

$\alpha = 0 \implies \|X\|_{\circledast,\alpha} = \|X\|_*$ (convex assuming known variance)

$\alpha > 0, \lambda_r \to \infty \implies \hat{X} \implies SVP(X,\alpha) = U_\alpha \Sigma_\alpha V_\alpha^T$ (nonconvex)

Experimental Setup

1. Define $W \in \mathbb{R}^{D \times N}$ to be a random Uniform matrix ($w_{ij} \sim U[0,1]$)
2. Knowing that SVD$(W) = U\Sigma V^T$, get $U_\alpha = U[:, 1:\alpha]$ for known $\alpha < D$
3. Let $\boldsymbol{z}_i \sim U[-a, a]$ or $\boldsymbol{z}_i \sim \mathcal{N}(0, a^2)$ such that $\boldsymbol{x}_i = U_\alpha \boldsymbol{z}_i \;\; \forall i$
4. Then, $\forall i \;\; \boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon}_i$ where $\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \nu_i)$
5. Collect all $\boldsymbol{y}_i$ such that $Y = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n]$ and $\Pi =$ diagm$(\nu_1, \ldots, \nu_n)$
6. Run opt. problem to find $\hat{X}$ and do SVD$(\hat{X}) \implies \hat{U}$
   Measure subspace affinity error $\|U_\alpha U_\alpha' - \hat{U}\hat{U}'\|_F / \|U_\alpha U_\alpha'\|_F$

Benchmark Comparisons

Probabilistic PCA (PPCA) [homoscedastic]

- Hard rank constraint, assumes same noise variance

Robust PCA (RPCA) [homoscedastic]

- RPCA is excluded from results to prevent straw man argument

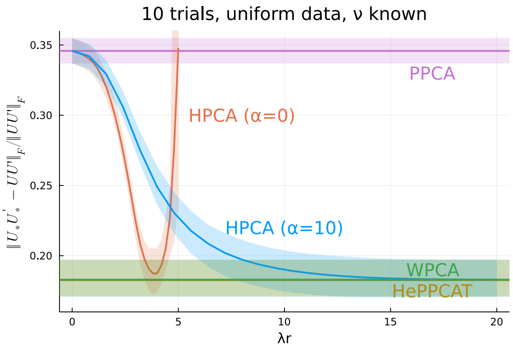
Weighted PCA (WPCA) [heteroscedastic]

- Weighted sample covariance method
- Hard rank constraint, assumes known variances

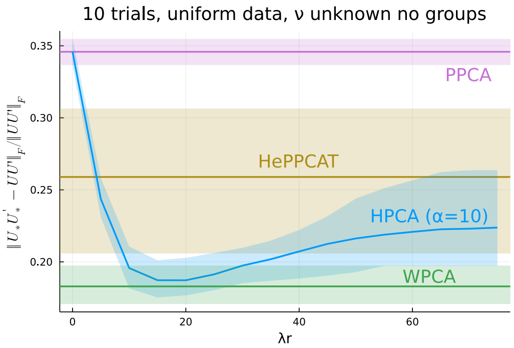
Heteroscedastic Probabilistic PCA Technique (HePPCAT) [hetero.]

- Factor analysis method
- Hard rank constraint, solves unknown variance case

Planted Model Results (Sample Heteroscedasticity)



known variance case

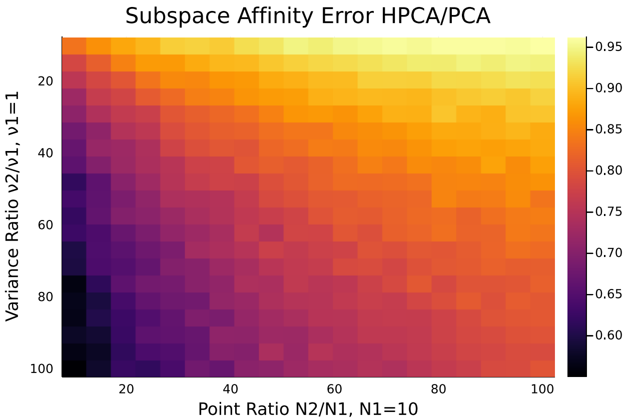


unknown variance case

ν_1	ν_2	Total Points	Ambient Dimension	Subspace Dimension	Good Samples
0.5	10	500	100	10	10

PCA Comparison

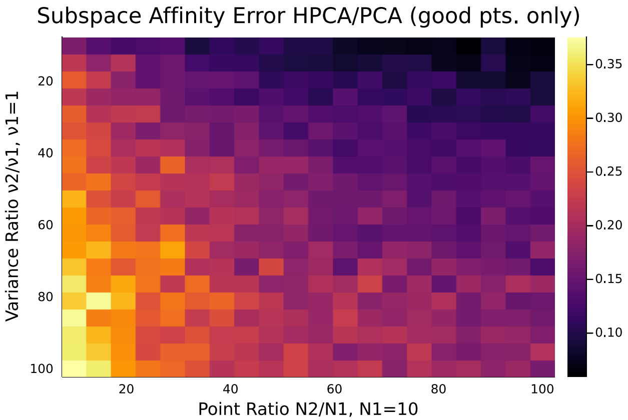
Param.	Value
Subspace Dim.	10
Ambient Dim.	100
Good Samples	10
ν_1	1
Basis Coordinate Range	100
Trials	10



known variance case

PCA Comparison (HPCA = All Pts., PCA = Good Pts.)

Param.	Value
Subspace Dim.	10
Ambient Dim.	100
Good Samples	10
ν_1	1
Basis Coordinate Range	100
Trials	10



known variance case

Algorithm Summary

- handles **both forms** of heteroscedasticity (not shown for brevity)
- no **distributional assumptions** about the low rank data
- **flexible** whether rank is known or not
- **convex** cases when the variances are known
- **extends** to linear operators (not shown for brevity)
- **outperforms** current methods like HeteroPCA and HePPCAT*

* at the cost of computation time & learning the regularization parameter

Future Work I

Fast HPCA (Burer-Monteiro decomposition + balancing preconditioners)

Let $X = UV^T$ where $U \in \mathbb{R}^{D \times k}$ and $V^T \in \mathbb{R}^{k \times N}$ for known rank k

$$\arg\min_{U,V,\Pi} \frac{1}{2} \| [Y - \underbrace{{\color{red}UV^T}}_{\text{nonconvex matrix factorization}}] \overbrace{\Pi^{-1/2}}^{\text{weighted noise}} \|_F^2 + \underbrace{\frac{D}{2} \log |\Pi|}_{\text{unknown variances}}$$

$$U_{k+1} = U_k - \alpha \nabla_U f(U_k, V_k, \Pi_k) \underbrace{(V_k^T V_k)^{-1}}_{\text{balancing term}} \qquad V_{k+1} = V_k - \alpha \nabla_V f(U_k, V_k, \Pi_k) \underbrace{(U_k^T U_k)^{-1}}_{\text{balancing term}}$$

Idea inspired by Yuejie Chi's work on ScaledGD [6]

Conclusion

- Heteroscedasticity is problematic for many algorithms
- Noisy data still contains information about the basis
- Robust PCA works for outliers but fails under heteroscedastic cases
- HPCA builds off Robust PCA for heteroscedastic situations
- Future work on fast implementation & Union of Subspaces (UoS)

Scratch Slide

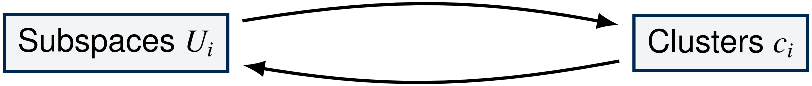
Future Work II

Heteroscedastic k-Subspaces

$$\min_{\mathcal{C},\mathcal{U}} \sum_{k=1}^{K} \sum_{i:y_i \in c_k} \|y_i - U_k U_k^T y_i\|_2^2$$

$\mathcal{C} = \{c_1, \ldots, c_k\}$ = estimated clusters, $\mathcal{U} = \{U_1, \ldots, U_K\}$ = subspace bases

$$\min_i \underbrace{\|\boldsymbol{y} - \overbrace{U_i U_i^T}^{\text{subspace projector}} \boldsymbol{y}\|}_{\text{residual}} \ \forall \boldsymbol{y} \in Y$$



$$\text{HPCA}(X_i \subset Y) \ \forall X_i \in c_i$$

Benchmark Comparison Algorithms I

Probabilistic PCA (PPCA)

In factor analysis, the following latent variable model relates observation data t to unobserved variable x:

$$t = Wx + \mu + \epsilon, \text{ where } x \sim \mathcal{N}(0, I), \epsilon \sim \mathcal{N}(0, \sigma^2 I), \; t \sim \mathcal{N}(\mu, \underbrace{WW^T + \sigma^2 I}_{C})$$

Then, for sample covariance matrix S_t, the log-likelihood is:

$$\mathcal{L} = -\frac{N}{2}(d \log(2\pi) + \log(\det(C)) + \mathrm{tr}(C^{-1}S))$$

The subspace basis is found by orthogonalizing W via SVD

Benchmark Comparison Algorithms II

Weighted PCA (WPCA)

Given weights $w_1, \ldots, w_N$ for points $y_1, \ldots, y_N$
form the weighted sample covariance matrix as:

$$S = \sum_{i=1}^{N} w_i [y_i y_i']$$

and get subspace basis by performing:

$$\hat{U} = [\hat{u}_1, \ldots, \hat{u}_k] = \text{EVD}(S)$$

for given rank k and weights $w_i = \frac{1}{\sigma_i^2}$ (a natural choice)

Benchmark Comparison Algorithms III

Heteroscedastic Probabilistic PCA Technique (HePPCAT)

For $n_1 + \ldots + n_L = n$ data samples from L noise groups, the model is described as

$$y_{l,i} = Fz_{l,i} + \epsilon_{l,i} \quad i \in \{1, \ldots, n_L\},\ l \in \{1, \ldots, L\}$$

for scores $z_{l,i} \sim \mathcal{N}(0, v_l I)$ and points $y_{l,i} \sim \mathcal{N}(0, FF^T + v_l I)$

Form the log-likelihood as the following:

$$\mathcal{L}(F, v) = \frac{1}{2} \sum_{l=1}^{L} [n_l \ln \det(FF^T + v_l I)^{-1} - \operatorname{tr}\{Y_l^T (FF^T + v_l I)^{-1} Y_l\}]$$

Feature HPCA

Let $Y = [\,\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\,] \in \mathbb{R}^{D\times N}$ and $W = \text{diagm}(w_1, \ldots, w_D) \in \mathbb{R}^{D\times D}$

Consider a model with heteroscedasticity across the features

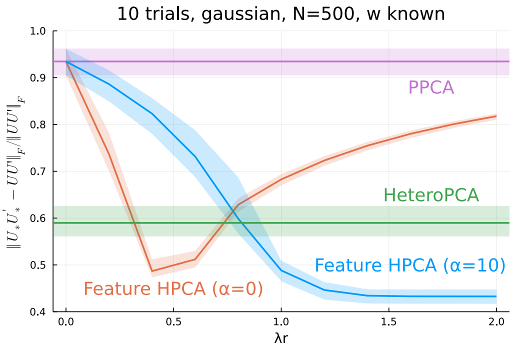
$$\boldsymbol{y}_i = \boldsymbol{x}_i + \boldsymbol{\epsilon} \text{ s.t. } \underbrace{\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, W)}_{\text{same for each point!}}$$

$$\operatorname*{arg\,min}_{X,Z,W} \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{\text{low rank}} + \underbrace{\frac{1}{2}\|W^{-1/2}Z\|_F^2}_{\text{weighted features}} + \underbrace{\frac{D}{2}\log|W|}_{\text{unk. quality}} \text{ s.t. } Y = X + Z$$

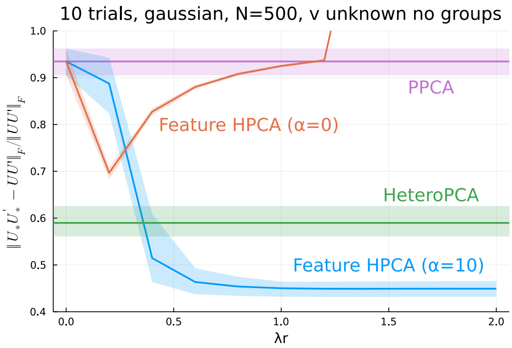
Note!

W updates $\implies$ using feature space sample covariance matrix

Planted Model Results (Feature Heteroscedasticity)



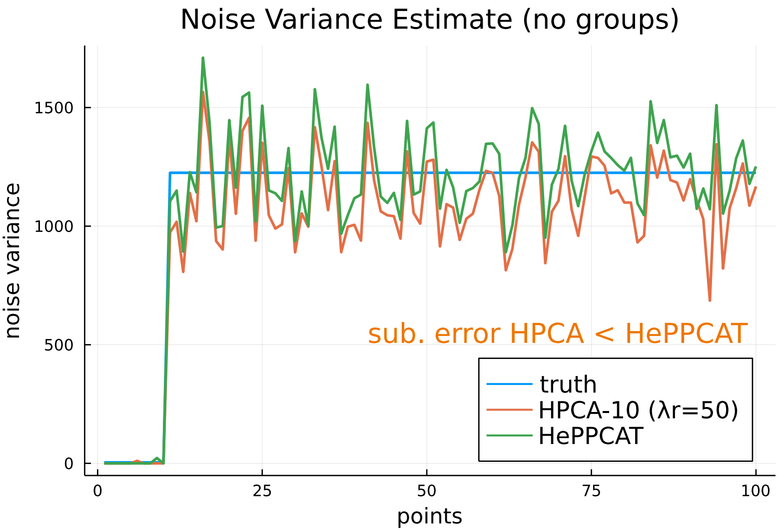
known variance case



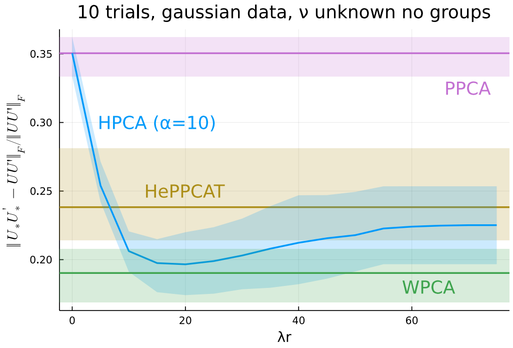
unknown variance case

w_1	w_2	Total Points	Ambient Dimension	Subspace Dimension	Good Features
2	35	500	100	10	50

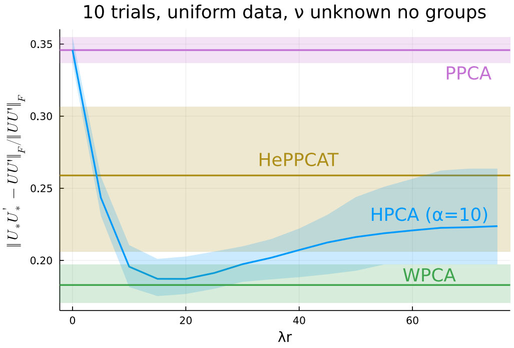
HPCA Variance Estimation



HPCA/HePPCAT Distribution Effects



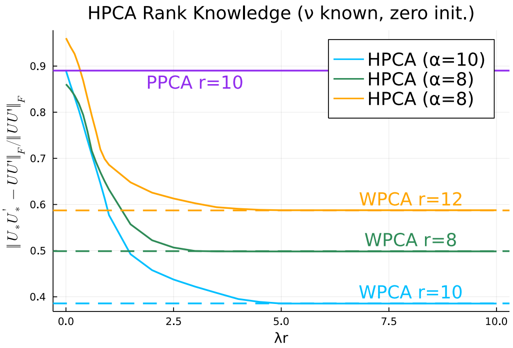
Gaussian case



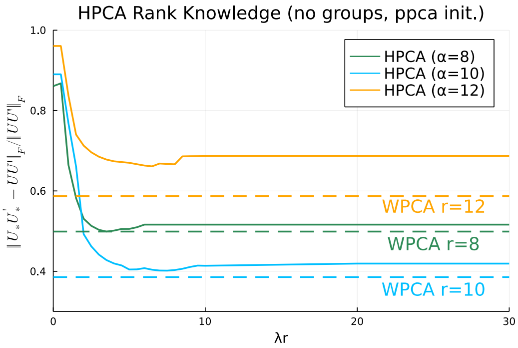
Uniform case

ν_1	ν_2	Total Points	Ambient Dimension	Subspace Dimension	Good Samples
0.5	10	500	100	10	10

HPCA Rank Knowledge



Known variance case



Unknown variance case

Low Rank Matrix Completion I

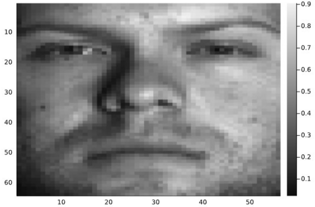
Consider a general linear mapping $\mathcal{A}(\cdot) : \mathbb{R}^{D\times N} \to \mathbb{R}^{D\times N}$

$$\arg\min_{X,\Pi} \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{\text{low rank}} + \underbrace{\frac{1}{2}\|[Y - \mathcal{A}(X)]\Pi^{-1/2}\|_F^2}_{\text{weighted noise}} + \underbrace{\frac{D}{2}\log|\Pi|}_{\text{unk. variance}}$$

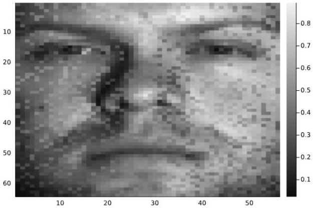
Examples

$\mathcal{A}(X) = M \odot X = \tilde{M}X$ (missing data i.e. matrix completion)
$\mathcal{A}(X) = H \circledast X_i \quad \forall i = \tilde{H}X$ (deconvolution e.g. phase retrieval)
$\mathcal{A}(X) = FX$ (forward models e.g. MRI encoding matrix)

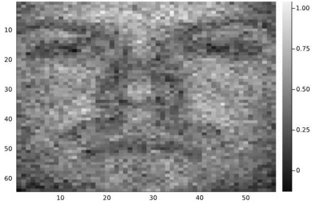
original image



low noise image

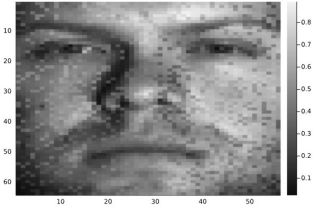


high noise image

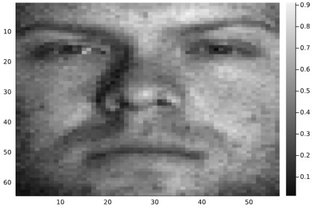


$\nu_1^{1/2}$	$\nu_2^{1/2}$	Total Samples	Good Samples	Ambient Dimension	Subspace Dimension	Corruption Factor
0.01	0.1	22	1	3584	≈ 4	20%

low noise image

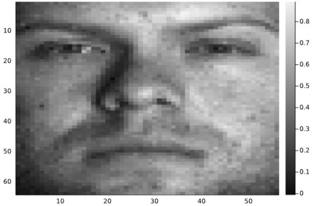


WPCA



PSNR: 29.2dB

HPCA



PSNR: 35.3dB

$\nu_1^{1/2}$	$\nu_2^{1/2}$	Total Samples	Good Samples	Ambient Dimension	Subspace Dimension	Corruption Factor
0.01	0.1	22	1	3584	≈ 4	20%

ADMM Implementation I

$$\underset{X,Z,\Pi}{\arg\min}\ \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{\text{low rank}} + \underbrace{\frac{1}{2}\|Z\Pi^{-1/2}\|_F^2}_{\text{weighted noise}} + \underbrace{\frac{D}{2}\log|\Pi|}_{\text{unk. variance}} \quad \text{s.t.} \quad Y = X + Z$$

Form the augmented Lagrangian function with penalty parameter μ:

$$\mathcal{L}_\mu(X, Z, \Lambda) = \lambda_r \|X\|_{\circledast,\alpha} + \frac{1}{2}\|Z\Pi^{-1/2}\|_F^2 + \langle \Lambda, Y - X - Z \rangle + \frac{\mu}{2}\|Y - X - Z\|_F^2 + \frac{D}{2}\log|\Pi|$$

Solve for each primal and dual individually X, Z, Λ, Π

ADMM Implementation II

X update:

$$X_{k+1} = \arg\min_{X_k} \mathcal{L}_\mu(X_k, Z_k, \Lambda_k) = \arg\min_{X_k} \lambda_r \|X_k\|_{\circledast,\alpha} + \frac{\mu}{2}\|Y - X_k - Z_k + \frac{1}{\mu}\Lambda_k\|_F^2$$

$$= \operatorname{prox}_{\lambda_r \mu^{-1}\|\cdot\|_{\circledast,\alpha}}(Y - Z_k + \frac{1}{\mu}\Lambda_k) = \text{PSSV}(Y - Z_k + \frac{1}{\mu}\Lambda_k, \frac{\lambda_r}{\mu})$$

PSSV - Notation

$$\text{PSSV}(A, \tau, \alpha) = U_A(D_{A1} + \mathcal{S}_\tau[D_{A2}])V_A^T$$

$$D_{A1} = \text{diag}(\sigma_1(A), \ldots, \sigma_\alpha(A), 0, \ldots, 0)$$

$$D_{A2} = \text{diag}(0, \ldots, 0, \sigma_{\alpha+1}(A), \ldots, \sigma_N(A))$$

$$\mathcal{S}_\tau[x] = \text{sign}(x)\max(|x| - \tau, 0)$$

ADMM Implementation III

Z update:

$$Z_{k+1} = \arg\min_{Z_k} \mathcal{L}_\mu(X_k, Z_k, \Lambda_k) = \arg\min_{Z_k} \frac{1}{2}\|Z_k\Pi^{-1/2}\|_F^2 + \frac{\mu}{2}\|Y - X_k - Z_k + \frac{1}{\mu}\Lambda_k\|_F^2$$

$$= [\mu(Y - X_k) + \Lambda_k](\Pi^{-1} + \mu I)^{-1}$$

Π update:

$$\Pi_{k+1} = \arg\min_{\Pi_k} \frac{1}{2}\|Z\Pi_k^{-1/2}\|_F^2 + \frac{D}{2}\log|\Pi_k| = \arg\min_{\Pi_k} \frac{1}{2}\operatorname{tr}(Z^T Z\Pi_k^{-1}) + \frac{D}{2}\log|\Pi_k|$$

$$\implies \nabla_{\Pi_k} f(\Pi_k) = \frac{-1}{2}(Z^T Z \odot \Pi_k^{-2}) + \frac{D}{2}\Pi_k^{-1}$$

$$\implies \Pi_{k+1} = \frac{1}{D}Z^T Z \odot I = \frac{1}{D}(Y - X)^T(Y - X) \odot I$$

HPCA Convergence

From [7], convergence is shown under the following ADMM framework

$$\min_{x,y} f(x) + g(y) \text{ s.t. } Ax + b = y$$

$$\text{W.L.O.G., HPCA} = \quad \arg\min_{X,Z} \underbrace{\lambda_r \|X\|_{\circledast,\alpha}}_{f(X)} + \underbrace{\frac{1}{2}\|Z\Pi^{-1/2}\|_F^2}_{g(Z)} \text{ s.t. } \underbrace{X}_{A=I} + Z = Y$$

1. $f(X)$: proper, lower semi-continuous, semi-algebraic function
2. $g(Z)$: continuous differentiable, semi-algebraic function w/ $L_{\nabla g} > 0$
3. $\exists \gamma \in \mathbb{R}$ s.t. $A^T A \succeq \gamma I$
4. $\{X_K, Z_k\}$ sequence generated by ADMM is bounded

1,2,3,4 $\implies$ convergence to KKT point (assuming aug. param. $\mu > 2L_{\nabla g}$)

References

[1] Y. Fu, W. Wang, and C. Wang, "Image change detection method based on rpca and low-rank decomposition," in *2016 35th Chinese Control Conference (CCC)*, IEEE, 2016, pp. 9412–9417.

[2] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic mri with separation of background and dynamic components," *Magnetic Resonance in Medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[3] Z. Zhou, X. Li, J. Wright, E. Candes, and Y. Ma, "Stable principal component pursuit," in *2010 IEEE international symposium on information theory*, IEEE, 2010, pp. 1518–1522.

References

[4] N. Vaswani, T. Bouwmans, S. Javed, and P. Narayanamurthy, “Robust subspace learning: Robust pca, robust subspace tracking, and robust subspace recovery,” *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 32–55, 2018.

[5] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim, and I. S. Kweon, “Partial sum minimization of singular values in robust pca: Algorithm and applications,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 4, pp. 744–758, 2015.

[6] T. Tong, C. Ma, and Y. Chi, *Accelerating ill-conditioned low-rank matrix estimation via scaled gradient descent*, 2020. DOI: 10.48550/ARXIV.2005.08898. [Online]. Available: https://arxiv.org/abs/2005.08898.

References

[7] K. Guo, D. Han, and T.-T. Wu, "Convergence of alternating direction method for minimizing sum of two nonconvex functions with linear constraints," *International Journal of Computer Mathematics*, vol. 94, no. 8, pp. 1653–1669, 2017.