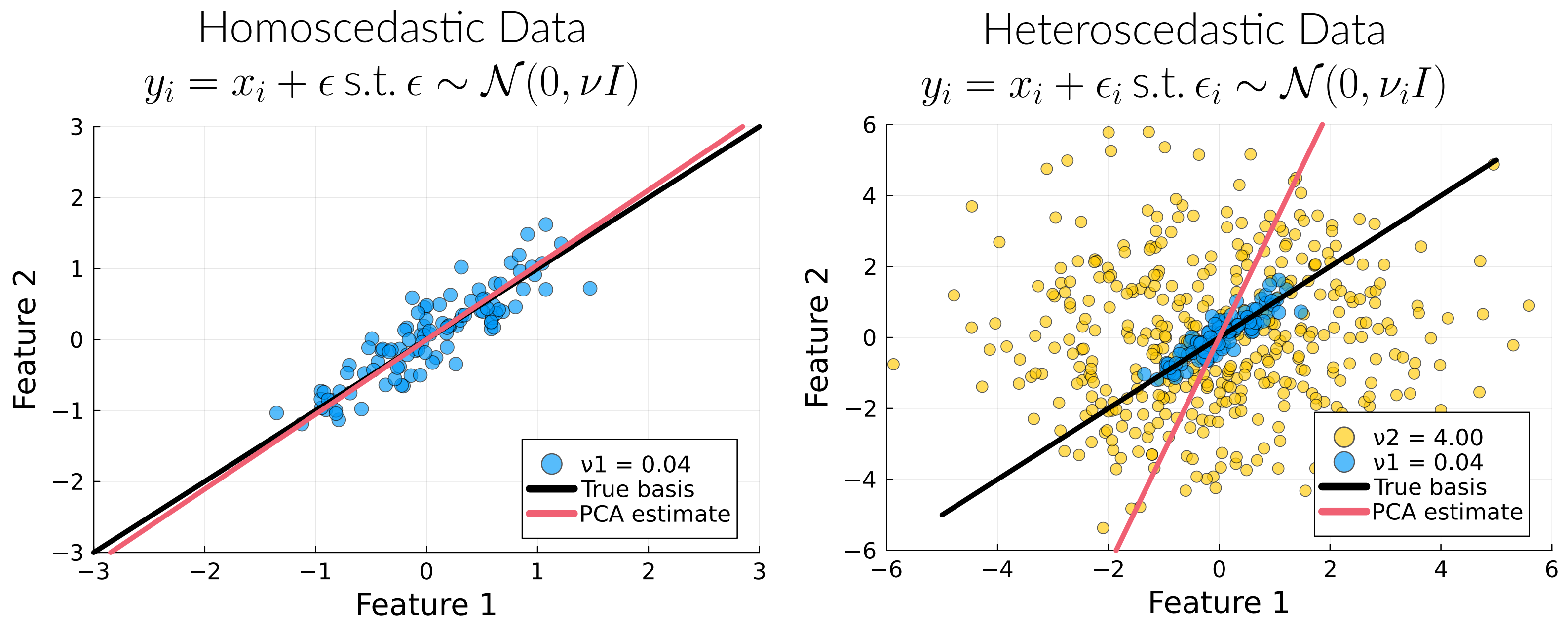
ALPCAH: Sample-wise Heteroscedastic PCA w/ Singular Value Regularization

Javier Salazar Cavazos, Jeffrey A. Fessler, Laura Balzano

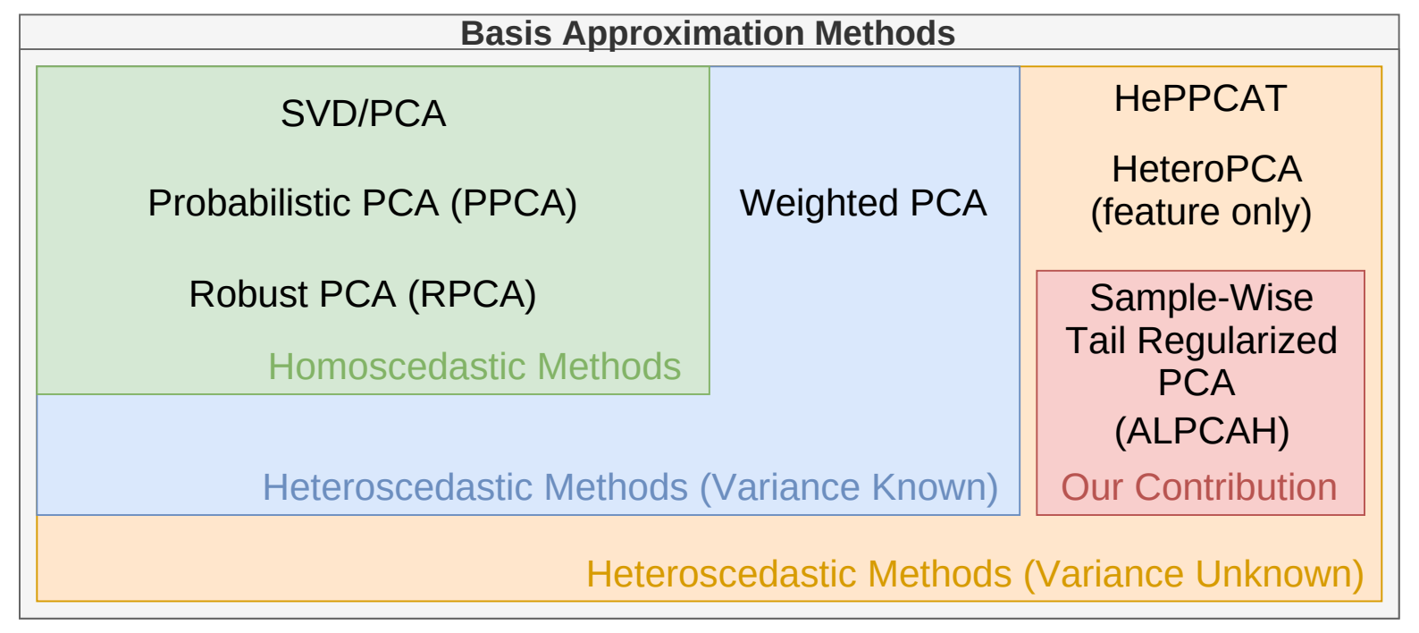
Electrical Engineering and Computer Science (EECS) Department, University of Michigan, Ann Arbor, Michigan, United States

Heteroscedasticity



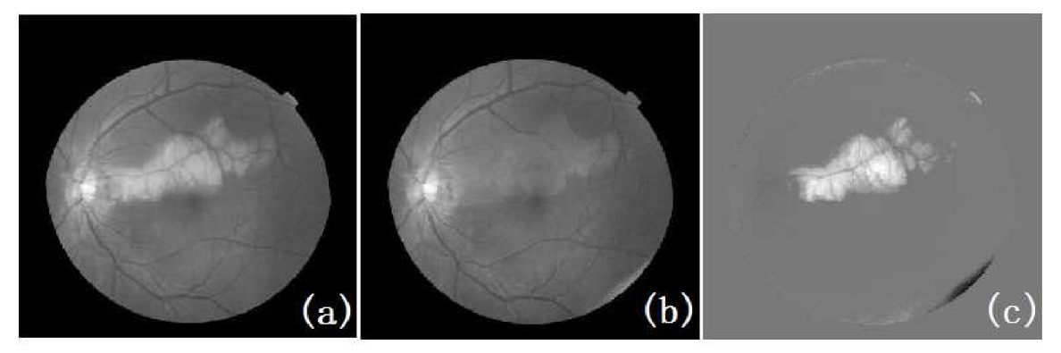
PCA methods like Robust PCA [2] and PCA work well in the homoscedastic setting, i.e., when the data is the same quality, but fail to accurately estimate the basis when the data varies in quality, i.e., in the heteroscedastic setting.

Background

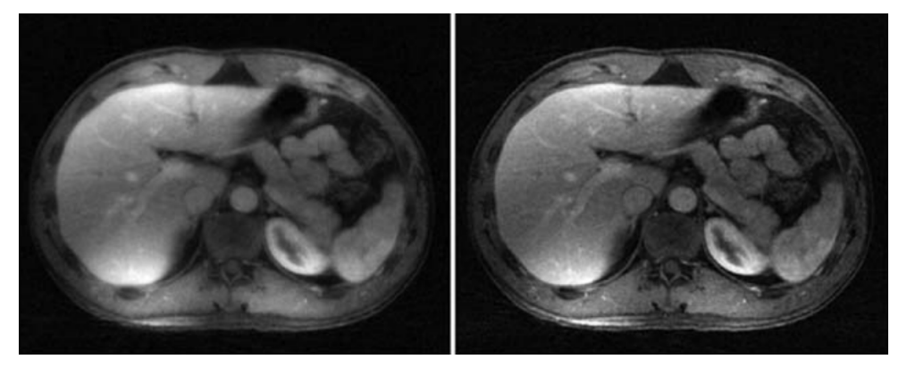

Basis Approximation Methods
SVD/PCA
Probabilistic PCA (PPCA)
Robust PCA (RPCA)
Homoscedastic Methods
Weighted PCA
Heteroscedastic Methods (Variance Known)
HePPCAT
HeteroPCA (feature only)
Sample-Wise Tail Regularized PCA (ALPCAH)
Our Contribution
Heteroscedastic Methods (Variance Unknown)

This paper develops a PCA method named ALPCAH that can estimate the sample-wise noise variances and use this information in the model to improve the estimate of the subspace basis associated with the low-rank structure of the data. This is done without distributional assumptions of the low-rank component and without assuming the noise variances are known which are some of the limitations of current methods like HePPCAT [4] and Weighted PCA [3] respectively.

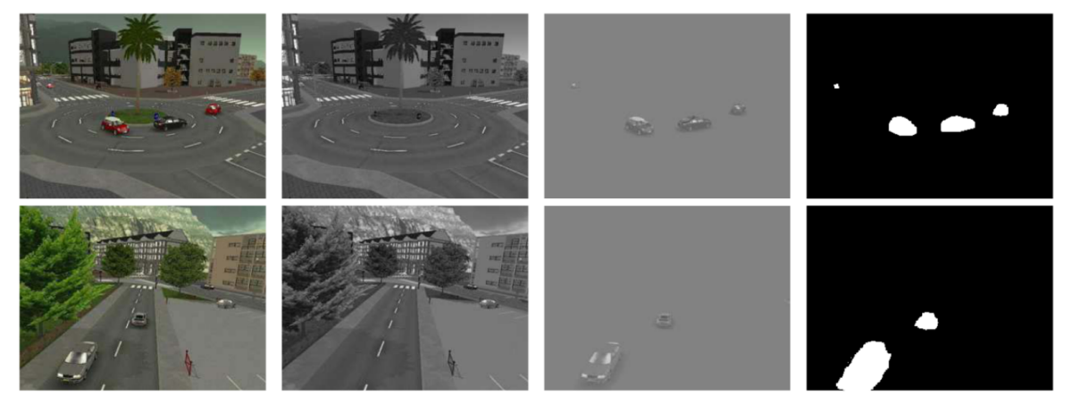
Applications



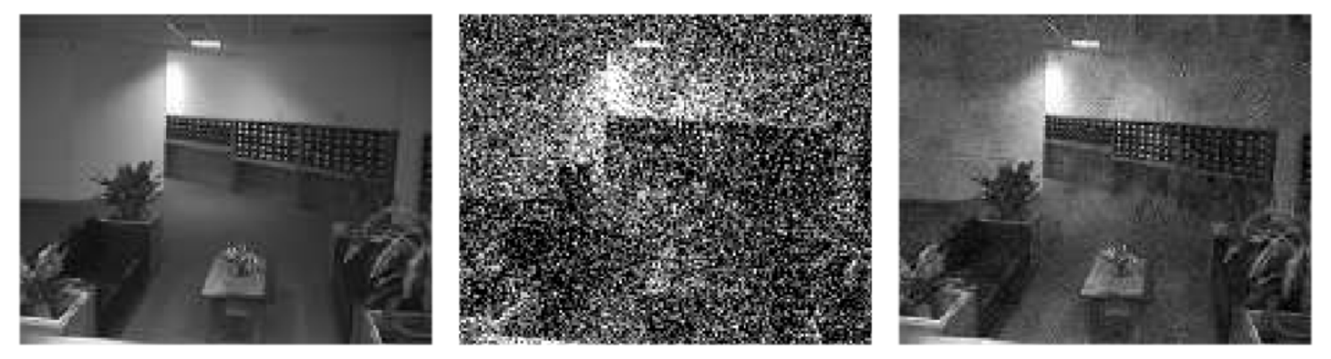
Lesion Detection [1]



Dynamic MRI Reconstruction [1]



Motion Estimation [1]



Image/Video Denoising [1]

Many modern data science problems require learning an approximate subspace basis for some data. This is useful for key tasks like dimensionality reduction.

References

[1] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, and R. Otazo. On the applications of robust pca in image and video processing. *Proceedings of the IEEE*, 106(8):1427–1457, 2018.

[2] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *CoRR*, abs/0912.3599, 2009.

[3] L. Delchambre. Weighted principal component analysis: a weighted covariance eigendecomposition approach. *Monthly Notices of the Royal Astronomical Society*, 446(4):3545–3555, 2015.

[4] D. Hong, K. Gilman, L. Balzano, and J. A. Fessler. HePPCAT: Probabilistic PCA for data with heteroscedastic noise. *IEEE Transactions on Signal Processing*, 69:4819–4834, 2021.

[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*, 38(4):744–758, 2015.

ALPCAH
Algorithm for Low-rank PCA for Heteroscedastic data

Let $y_i \in \mathbb{R}^D$ represent the data samples for index $i \in \{1, \dots, N\}$ given N total samples and D represent the ambient dimension. Let x_i represent the low-dimensional data sample generated by $x_i = Uz_i$ where $U \in \mathbb{R}^{D \times k}$ is an orthogonal basis of dimension k and $z_i \in \mathbb{R}^k$ are basis coordinates. Then the heteroscedastic model is described as follows assuming Gaussian noise:

$$y_i = x_i + \epsilon_i \text{ where } \epsilon_i \sim \mathcal{N}(0, \nu_i I) \tag{1}$$

for noise variances ν_i. For the measurement model $y_i \sim \mathcal{N}(x_i, \nu_i I)$, the probability density function for a single point is

$$\frac{1}{\sqrt{(2\pi)^k |\nu_i I|}} \exp\left[-\frac{1}{2}(y_i - x_i)^T (\nu_i I)^{-1} (y_i - x_i)\right]. \tag{2}$$

For uncorrelated samples, the joint log likelihood of all y_i is the following after dropping constants

$$\sum_{i=1}^{N} -\frac{1}{2}\log|\nu_i I| - \frac{1}{2}(y_i - x_i)^T (\nu_i I)^{-1} (y_i - x_i). \tag{3}$$

Let $\Pi = \text{diag}(\nu_1, \dots, \nu_N) \in \mathbb{R}^{N \times N}$ be a diagonal matrix representing the (typically unknown) noise variances. Let $Y = [y_1, \dots, y_N] \in \mathbb{R}^{D \times N}$ represent all of the data samples. Then, the log likelihood in matrix form is

$$-\frac{D}{2}\log|\Pi| - \frac{1}{2}\text{Trace}[(Y - X)^T \Pi^{-1} (Y - X)]. \tag{4}$$

Using trace properties, the optimization problem we pose for the heteroscedastic model is

$$\underset{X,\Pi}{\arg\min}\, \lambda f_k(X) + \frac{1}{2}\|(Y - X)\Pi^{-1/2}\|_F^2 + \frac{D}{2}\log \underbrace{|\Pi|}_{\text{determinant}} \tag{5}$$

where $f_k(X)$ is a relatively new functional in the literature [5] that promotes low-rank structure in X by penalizing the tail singular values:

$$f_k(X) \triangleq \sum_{i=k+1}^{\min(D,N)} \sigma_i(X) = \|X\|_* - \|X\|_{\text{Ky-Fan}(k)}, \tag{6}$$

where $\sigma_i(X)$ is the ith singular value of X, $\|\cdot\|_*$ is the nuclear norm, and $\|\cdot\|_{\text{Ky-Fan}(k)}$ is the Ky-Fan norm defined as the sum of the first k singular values. For $k = 0$, $f_0(X) = \|X\|_*$. For a general $k > 0$, $f_k(X)$ is a nonconvex difference of convex functions. When $k > 0$ and $\lambda \to \infty$, then the solution of the optimization problem approaches $\hat{X} = \sum_{i=1}^{k} \sigma_i u_i v_i' \in \mathbb{R}^{D \times k}$ meaning the solution becomes identical to a singular value projection approach.

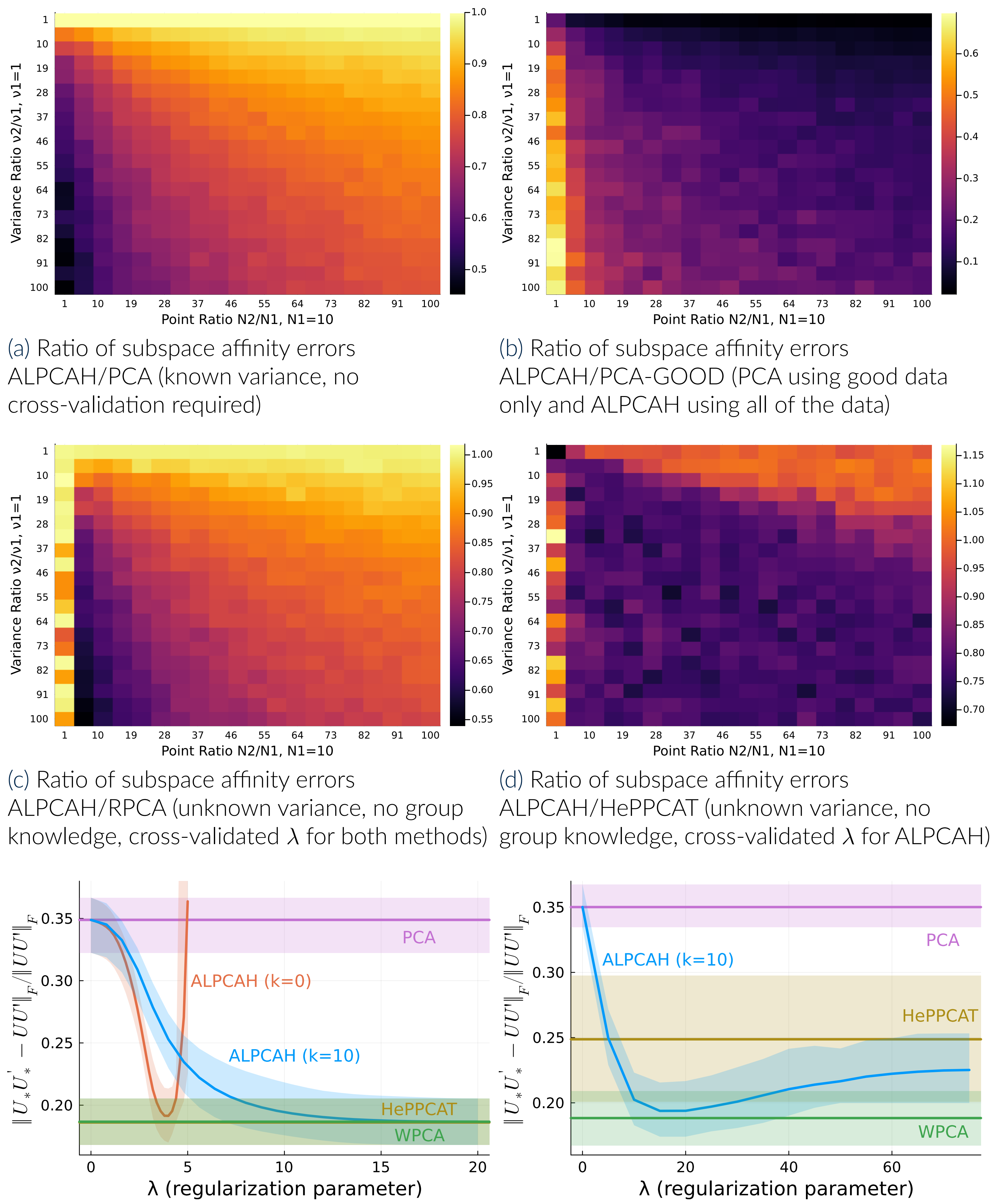
Simulation Study

Let $U \in \mathbb{R}^{100 \times 10}$ represent a 10 dimensional subspace generated by random uniform matrices such that $U\Sigma V^T = \text{svd}(A)$, where $A_{i,j} \sim \mathcal{U}[0, 1]$. The low-rank data x_i we simulated as $x_i = Uz_i$ where the coordinates $z_i \in \mathbb{R}^{10}$ were generated from $\mathcal{U}[-100, 100]$ for each element in the vector. Then, we generated $y_i = Uz_i + \epsilon_i$ where $\epsilon_i \in \mathbb{R}^{100}$ is drawn from $\mathcal{N}(0, \nu_i I)$. The noise variance for group 1 (ν_1) was fixed to 1 and we varied group 2 noise variances (ν_2). The error metric used is subspace affinity error that compares the difference in projection matrices $\|UU' - \hat{U}\hat{U}'\|_F / \|UU'\|_F$ so that a low error signifies a closer estimate of the true subspace.

Funding Disclosure

This work is supported in part by NSF CAREER Grant 1845076.

Simulation Results



(a) Ratio of subspace affinity errors ALPCAH/PCA (known variance, no cross-validation required)

(b) Ratio of subspace affinity errors ALPCAH/PCA-GOOD (PCA using good data only and ALPCAH using all of the data)

(c) Ratio of subspace affinity errors ALPCAH/RPCA (unknown variance, no group knowledge, cross-validated λ for both methods)

(d) Ratio of subspace affinity errors ALPCAH/HePPCAT (unknown variance, no group knowledge, cross-validated λ for ALPCAH)

(e) Subspace affinity error of various PCA methods as the regularization parameter is adjusted (known variance)

(f) Subspace affinity error of various PCA methods as λ is adjusted (unknown variance, no group knowledge)

Figure 1. Heatmaps and plots of ALPCAH results on synthetic data to explore heteroscedasticity effects on subspace basis approximation

Key Findings

In the known variance case, Figure 1a shows that ALPCAH performs well relative to PCA in noisy situations and can improve estimation by up to 50% or 20% in more tame situations. From Figure 1b, clearly it is beneficial to collect and use all of the data, since the noisy points offer meaningful information that can improve the estimate of the basis versus using good data alone. For the unknown variance case, Figure 1d generally shows that on average there was a 20% improvement over HePPCAT. Since HePPCAT is a hard rank constraint method, it seems beneficial to not completely shrink the tail singular values but rather to retain them as they seem to improve the estimation process. Moreover, since we make no distributional assumptions about X itself besides low-rank assumptions, then this assumption relaxation helps us achieve lower error in settings where the basis coordinates are not Gaussian, whereas HePPCAT makes Gaussian assumptions about the basis coordinates themselves.