Robust Priors for Bayesian PCA under Randomized Projections

Thomas Cole (twc2131)

December 2025

Abstract

Randomized projections are a common tool for reducing the computational cost of high-dimensional models, but they introduce approximation noise that can distort latent factor recovery. This project investigates how common priors in Bayesian PCA—Gaussian, Student-t, Laplace, Log-quadratic, and Spike-and-Slab—respond to this projection-induced noise. Using synthetic datasets with known ground-truth factors and validation on the classic Olivetti Faces benchmark, we apply Gaussian randomized projections (with power iteration) followed by black-box variational inference for posterior estimation. We evaluate reconstruction error, factor recovery, and computational speedups across both dense and sparse regimes. We find that while projections preserve low reconstruction errors across all priors with dramatic speedups, they erase coordinate-wise sparsity patterns that distinguish priors in full-data settings. Spike-and-slab shows marginal robustness, suggesting projections excel for scalable dimensionality reduction but sacrifice interpretable factor loadings.

1 Introduction

Many modern datasets are high-dimensional, and randomized projections are a common tool for reducing computational burden before fitting statistical models. While these projections preserve global structure in expectation, they introduce projection-induced approximation noise that can distort local structure, precisely the information latent-factor models rely on. Bayesian PCA offers a flexible, uncertainty-aware framework for factor analysis, but how different priors behave once data is compressed into a lower-dimensional subspace remains poorly understood.

Most prior work on Bayesian PCA assumes access to original data and focuses on priors' influence on shrinkage, sparsity, and identifiability. Less is known about fitting models downstream of randomized projections, where latent structure is only partially preserved and covariance eigenstructure is systematically perturbed. This gap motivates our question: how do different priors in Bayesian PCA respond to projection-induced distortions?

Our contributions include a systematic empirical comparison of common priors (Gaussian, Student-t, Laplace, Log-quadratic[†], Spike-and-Slab) under randomized projections using synthetic data with known ground truth, quantifying impacts on reconstruction and factor recovery, plus validation on the Olivetti Faces dataset with practical guidelines for prior selection in compressed settings.

[1] For Log-quadratic prior definition see Appendix A

2 Background

2.1 Bayesian PCA

As the name suggests, Bayesian PCA [1], formulates the standard PCA model probabilistically. Given observed data

$$X \in \mathbf{R}^{N \times D}, \quad X = WZ + \mu + \epsilon, \quad Z \sim \mathbf{N}(0, I_K), \quad \epsilon \sim \mathbf{N}(0, \sigma^2 I_D),$$

where $W \in \mathbf{R}^{D \times K}$ is the loading matrix, $\mu \in \mathbf{R}^D$ is the mean, and $K \ll D$ is the latent dimensionality, the joint distribution factorizes as

$$p(X, Z, W, \sigma^2) = p(X \mid Z, W, \sigma^2)\, p(Z)\, p(W)\, p(\sigma^2).$$

As usual, priors on W play a central role: they regularize the factorization, influence identifiability, and control how aggressively small or noisy loadings are shrunk toward zero.

2.2 Randomized Projections

Randomized projections compress high-dimensional data $X \in \mathbf{R}^{N \times D}$ via $\widetilde{X} = XR$, where $R \in \mathbf{R}^{D \times m}$ ($m \ll D$) is a random matrix, typically Gaussian to satisfy Johnson–Lindenstrauss guarantees [2, 4]. While preserving pairwise distances in expectation, these introduce approximation noise that perturbs the covariance eigenstructure critical for PCA [7]. In our pipeline, we additionally improve subspace approximation using randomized power iteration [3].

2.2.1 Projection-induced distortion vs. observation noise

To motivate our study, we distinguish projection-induced distortion from observation noise, the typical setting for prior robustness. Figure 1 shows this on Olivetti Faces: Gaussian observation noise (middle) differs qualitatively from projection reconstruction artifacts (right), despite matched magnitudes. This contrast motivates whether loading matrix priors can absorb projection distortion alongside factorization regularization.

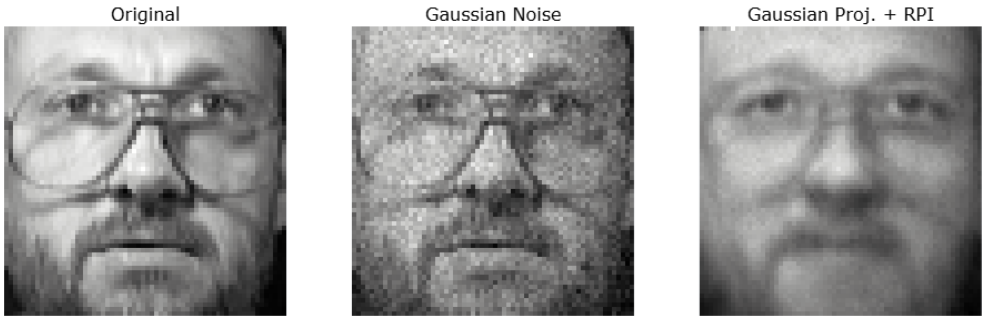

Figure 1: Original (left), Gaussian observation noise (middle), and Gaussian random projection reconstruction (right) on Olivetti Faces.

3 Methodology

3.1 Compression Stage

We first compress each $N \times D$ data matrix X into a lower-dimensional subspace using a Gaussian randomized projection with power iteration with a fixed target projection dimension. To improve the approximation of the dominant subspace, we perform one sequence of power iteration and set a fixed oversampling parameter of 10. The compression algorithm is given in Algorithm 4

3.2 Inference Procedure

To perform posterior inference, we employ black-box variational inference (BBVI) [5]. Since several of the priors we study are non-conjugate, deriving closed-form coordinate ascent updates is not feasible. BBVI provides a unified and modular inference scheme that only requires evaluating the log-joint density and its gradients, allowing us to reuse the same inference engine across all prior choices. To perform this, we posit a fully factorized Gaussian variational distribution over both the factor loadings W and latent factors Z, while treating σ^2 as fixed:

$$q(W, Z) = q(W)\, q(Z), \quad q(W) = \prod_{d,k} \mathbf{N}\Big(W_{dk} \mid \mu^W_{dk}, (\sigma^W_{dk})^2\Big), \quad q(Z) = \prod_{n,k} \mathbf{N}\Big(Z_{nk} \mid \mu^Z_{nk}, (\sigma^Z_{nk})^2\Big)$$

We use the reparameterization trick to obtain low-variance Monte Carlo estimates of ELBO. Then we maximize the ELBO with respect to all variational parameters using stochastic gradient ascent with the Adam optimizer. Gradients of the Monte Carlo estimator are obtained automatically by differentiating through the re-parameterized samples (Appendix B). After convergence, posterior mean estimates for the factors and loadings serve as the recovered structure under each prior. All implementations are coded in Python using PyTorch and NumPy, with full code available at: `https://github.com/thomaswcole/pmml-final-project`.

4 Experiments

We conduct experiments on both synthetic datasets and a classic dimensionality reduction benchmark. Our motivation is twofold: first, to assess how prior selection in Bayesian PCA interacts with randomized projections on controlled synthetic data with known latent factors; and second, to validate these insights on the Olivetti Faces dataset, providing an applied check.

To evaluate performance, we measure computational time, relative reconstruction error, and, for synthetic experiments, factor recovery accuracy. These metrics allow us to compare the effectiveness of different priors in mitigating projection noise while maintaining reliable inference.

4.1 Synthetic Experiments

4.1.1 Synthetic Data Generation

To evaluate Bayesian PCA under randomized projections, we generate synthetic datasets with known latent structure. Each dataset consists of an $N \times D$ matrix X constructed as

$$X = ZW^\top + \epsilon,$$

where $Z \in \mathbf{R}^{N \times r}$ contains the latent factors, $W \in \mathbf{R}^{D \times r}$ is the loading matrix, and $\epsilon \sim \mathbf{N}(0, \sigma^2 I_D)$ represents Gaussian noise. The latent dimension is fixed at $r = 20$. After adding noise, columns of X are centered to have zero mean. We consider both dense and sparse loading matrices, controlled by a sparsity parameter δ. The latent factors Z are always drawn from a dense Gaussian distribution.

Datasets are generated with N ranging from 1,000 to 4,000, and $D = N$ fixed, with $\delta \in \{0.01, 1.0\}$. This design allows us to evaluate further allows us to evaluate factor recovery, since the true latent factors Z and loadings W are known. A reference algorithm is given in Appendix C.

4.1.2 Experimental Setup

For each synthetic dataset, we fit Bayesian PCA with priors: Gaussian, Student-t, Laplace, Log-quadratic, Spike-and-Slab. We initialize BBVI with $q(W) \sim N(0, 1)$, $q(Z) \sim N(0, 1)$. We fix the latent dimension $K = 20$, learning rate 0.01, and max iterations of 2000. Projected experiments use randomized power iteration ($q = 20$, oversampling=10). All run 10 times and we report averaged results. Our evaluated metrics include:

- Reconstruction: $\frac{\|X - \hat{X}\|_F}{\|X\|_F}$
- Factor recovery: $\frac{1}{K} \sum_k |\text{corr}(W_{\text{true},:k}, \hat{W}_{:k})|$
- Runtime: wall-clock time (10 seeds)

4.1.3 Results

Our results show that Bayesian PCA with randomized projections matches full-data reconstruction performance across all priors on synthetic datasets with $N = D$. Relative reconstruction errors remain low in the dense regime (~ 0.003) and comparable sparse regime (~ 0.024), consistently achieving similar performance to their un-projected baselines across all N (Tables 1 and 6). Most priors show near-identical projected vs non-projected errors, with spike-and-slab exhibiting slightly higher variance but maintaining competitive performance.

Factor recovery metrics indicate that projected inference achieves comparable subspace alignment ($\sim 0.17 - 0.20$) across priors and regimes, with spike-and-slab yielding marginally highest scores (Tables 3 and 8). Prior differences remain minimal in posterior factor recovery, which we investigate this further in the Olivetti Faces experiment.

As expected, randomized projections deliver substantial speedups that grow super-linearly with dimension: $\sim 3\times$ faster at $N = 1000$ and scaling to $\sim 50\times$ at $N = 4000$, with consistent patterns

across sparse regimes (Tables 2 and 7). These gains are a result of reduced computational complexity in the projected subspace for matrix operations and likelihood evaluations, which enables scalable Bayesian inference without compromising reconstruction accuracy or factor recovery.

Table 1: Reconstruction error (dense regime) across priors and sample sizes $N = D$.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	0.0033	0.0053	0.0030	0.0051	0.0029	0.0051
Student-t	0.0033	0.0053	0.0030	0.0051	0.0029	0.0051
Laplace	0.0033	0.0053	0.0030	0.0051	0.0029	0.0051
Log-quadratic	0.0033	0.0053	0.0030	0.0051	0.0029	0.0051
Spike-and-Slab	0.0034	0.0053	0.0030	0.0051	0.0029	0.0051

Table 2: Runtime (dense regime) across priors and sample sizes $N = D$, comparing projected vs non-projected.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	1.60	4.76	1.88	21.8	2.55	124
Student-t	1.57	4.77	1.94	22.1	2.60	124
Laplace	1.51	4.66	1.89	21.8	2.55	124
Log-quadratic	1.53	4.71	1.92	22.0	2.56	124
Spike-and-Slab	1.79	5.20	2.17	22.9	2.84	126

Table 3: Factor recovery (dense regime) across priors and sample sizes $N = D$, comparing projected vs non-projected.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	0.197	0.185	0.183	0.163	0.189	0.175
Student-t	0.197	0.185	0.183	0.163	0.189	0.175
Laplace	0.197	0.185	0.183	0.163	0.189	0.175
Log-quadratic	0.197	0.185	0.183	0.163	0.189	0.175
Spike-and-Slab	0.198	0.179	0.185	0.162	0.190	0.174

4.2 Olivetti Faces Experiment

4.2.1 Dataset and Preprocessing

We evaluate our method on the Olivetti Faces dataset [6], which contains 400 grayscale images of 40 individuals at 64×64 pixels. Each image is flattened into a 4096-dimensional vector, and these vectors are stacked into a data matrix $X \in \mathbf{R}^{400\times4096}$. No additional pre-processing was applied.

4.2.2 Experimental Setup

We again fit Bayesian PCA with priors: Gaussian, Student-t, Laplace, Log-quadratic, Spike-and-Slab. BBVI is initialized with $q(W), q(Z) \sim N(0, 1)$ and uses the same settings as the synthetic experiments: latent dimension $K \in \{10, 20, 30\}$, learning rate 0.01, and 2000 iterations. For projected experiments, we compress to $q = K$ dimensions using randomized power iteration (1 iteration, oversampling=10). Each (prior, K) configuration is run 10 times with different random seeds and results are averaged. We report the same metrics as the synthetic experiments: reconstruction error, final ELBO, and wall-clock time, using non-projected Bayesian PCA as baseline.

4.2.3 Results

Tables 4 and 5 summarize reconstruction error and computation time across priors and $K \in \{10, 20, 30\}$. Randomized projection introduces only a small accuracy loss (∼0.002 increase in relative error), with all priors performing within a narrow band and error decreasing with K. Spike-and-slab shows slightly lower error at higher K, though differences remain small. Notably, spike-and-slab exhibits the smallest projection gaps across all K (e.g., 0.0013 at $K = 20$ vs 0.0023 for Log-quadratic), confirming its robustness to projection-induced distortion. Computationally, projection yields a consistent ∼10× speedup (1s vs 10–14s per run), independent of prior choice and driven by dimensionality reduction from $D = 4096$ to $q = K$.

One limitation of randomized projections, not captured in our synthetic experiments, is their effect on factor interpretability. Figure 3 shows that projected models produce loading weight distributions that are nearly indistinguishable across priors (except spike-and-slab), while full-space models preserve distinct prior-induced sparsity patterns. This indicates that dense projections destroy the original coordinate structure where coordinate-wise priors are meaningful.

Table 4: Reconstruction error across priors and latent dimensions K, comparing projected vs non-projected data.

Prior	K=10		K=20		K=30	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	0.1462	0.1441	0.1227	0.1208	0.1112	0.1081
Student-t	0.1464	0.1440	0.1233	0.1208	0.1111	0.1079
Laplace	0.1462	0.1440	0.1231	0.1207	0.1110	0.1079
Log-quadratic	0.1460	0.1440	0.1233	0.1210	0.1110	0.1078
Spike-and-Slab	0.1459	0.1450	0.1227	0.1240	0.1108	0.1135

For model criticism, we select $K = 20$ based on lowest average ELBO across priors (Table 9). We treat reconstructed faces as approximate posterior predictive draws and visually inspect the resulting prior-dependent artifacts (Figure 2). We find that Gaussian and Student-t priors produce blurry faces, while sparsity-inducing priors (Laplace, Log-quadratic) reduce noise but over-smooth details and finally, Spike-and-Slab generates the sharpest features with minimal distortion.

Table 5: Average computation time (seconds) across priors and latent dimensions K, comparing projected vs non-projected data.

Prior	K=10		K=20		K=30	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	1.190	12.375	1.183	13.280	1.454	14.452
Student-t	1.337	14.221	1.411	13.816	1.317	14.714
Laplace	1.054	12.846	1.112	11.423	1.252	14.392
Log-quadratic	1.194	12.959	1.253	12.290	1.271	12.279
Spike-and-Slab	1.277	13.189	1.519	13.926	1.560	14.050

5 Conclusion

In this work, we examined how randomized projections interact with Bayesian PCA across a range of common priors, implemented using Black-Box Variational Inference. On synthetic low-rank data, projection preserved reconstruction and made different priors behave similarly, suggesting that in simple settings the projection step is largely benign. On faces, however, full-space models with a Spike-and-Slab prior produced relatively sparser loadings, while projected models yielded back-projected loadings whose distributions were nearly identical across priors, indicating that dense projections destroy the coordinate system in which these priors are interpretable.

This project suggests that, while Bayesian PCA with randomized projections can in fact perform well as its full-space Bayesian PCA counterpart, the gain in speed comes at the loss of factor interpretation, which, if that is the goal, would suggest using other methods. However, if the goal of Bayesian PCA is solely performance, including uncertainty metrics unlike classical PCA, then projections deliver massive speedups while preserving those metrics, but at the cost of nullifying prior impacts on sparsity and interpretability; under our setup, prior choice thus becomes largely irrelevant.

5.1 Additional approaches

As we explored, projections nullify differences across priors, so we tested two strategies to recover sparsity patterns post-projection. First, we re-parameterized loadings as $W = B\Theta$ using a patch-based dictionary B and sparsity priors on Θ; this worsened reconstruction error without restoring sparsity. Secondly, we tried jointly optimizing the projection matrix Q with variational parameters, but this yielded degenerate solutions and poor performance. These negative results show that simple changes cannot recover coordinate-wise sparsity, leaving projection-compatible sparse priors as an open challenge.

A Priors

We summarize below the un-normalized densities for each prior used on the loading matrix W.

Gaussian prior.

$$p(W) \propto \exp\left(-\tfrac{1}{2}\sum_{ij} W_{ij}^2\right).$$

Laplace prior.

$$p(W) \propto \exp\left(-\sum_{ij} |W_{ij}|\right).$$

Smooth log-quadratic prior

$$p(W) \propto \prod_{ij} \frac{1}{1+W_{ij}^2}.$$

Student-t prior ($\nu = 3$).

$$p(W) \propto \prod_{ij} \left(1 + \frac{W_{ij}^2}{\nu}\right)^{-(\nu+1)/2}.$$

Spike–slab prior ($\pi = 0.3$).

$$p(W_{ij}) = \pi\,\mathbf{N}(0,\sigma_1^2) + (1-\pi)\,\mathbf{N}(0,\sigma_0^2), \qquad p(W) \propto \prod_{ij}\left[\pi\, e^{-W_{ij}^2/(2\sigma_1^2)} + (1-\pi)\, e^{-W_{ij}^2/(2\sigma_0^2)}\right].$$

B BBVI Implementation Reference

Algorithm 1 BBVI Optimization for Bayesian PCA

Require:

- Data $Y \in \mathbf{R}^{N\times q}$
- Latent dimension K
- Prior type (Gaussian, Student-t, etc.)
- BBVI settings: steps=2000, lr=0.01

1: Initialize mean-field $q(W,Z) = \prod \mathbf{N}(\mu_W, e^{2\log\sigma_W})$
2: **for** $t = 1,\ldots,2000$ **do**
3: elbo $\leftarrow$ ELBO-MC(Y) *Alg. 2*
4: $\theta \leftarrow \text{Adam}(\theta, \nabla_\theta \text{elbo})$
5: **end for**
6: **return** $\bar{Z} = \mu_Z, \bar{W} = \mu_W$

Algorithm 2 MC ELBO Estimate

Require:

- Data $Y \in \mathbf{R}^{N \times q}$
- Variational params $\theta = \{\mu_W, \log \sigma_W, \mu_Z, \log \sigma_Z\}$
- Monte Carlo samples S

$\text{elbo}_{\text{MC}} \leftarrow 0$
for $s = 1, \ldots, S$ **do**
  $\tilde{W} \sim \text{explog}(\mu_W, \log \sigma_W)$
  $\tilde{Z} \sim \text{explog}(\mu_Z, \log \sigma_Z)$
  $\log p(Y|\tilde{Z}, \tilde{W}) = -\frac{1}{2\sigma^2}\|Y - \tilde{Z}\tilde{W}^\top\|^2$
  $\log p(\tilde{W}, \tilde{Z})$ *priors Appendix A*
  $\text{elbo}_{\text{MC}} \leftarrow \text{elbo}_{\text{MC}} + \mathbf{E}_q[\log p - \log q]$
end for
return $\text{elbo}_{\text{MC}}/S$

C Data Generation Algorithm

Algorithm 3 Synthetic Data Generation

Require:

- Dataset size N, ambient dimension D
- Latent dimension $r = 20$
- Noise level $\sigma = 0.01$
- Sparsity $\delta \in \{0.01, 1.0\}$

$Z_{\text{true}} \sim \mathbf{N}(0, I_r)^N$
$W_{\text{true}} \sim \mathbf{N}(0, 1)^{D \times r}$
if $\delta < 1$ **then**
  Sparsify W_{true}: keep each entry w.p. δ
end if
$X \leftarrow Z_{\text{true}} W_{\text{true}}^\top + \sigma \mathbf{N}(0, I_D)^N$
Center columns of X
return $X, Z_{\text{true}}, W_{\text{true}}$

D Randomized Power Iteration

The compression algorithm we utilize, via power iteration is given below.

Algorithm 4 Randomized Compression via Power Iteration

Require:

- Data matrix $X \in \mathbf{R}^{N \times D}$
- Target dimension q
- Number of power iterations n_{iter}
- Oversampling parameter p

$l \leftarrow q + p$
Generate Gaussian random matrix $G \in \mathbf{R}^{D \times l}$
$Y \leftarrow XG$
for $i = 1$ **to** n_{iter} **do**
    $Y \leftarrow X(X^\top Y)$
end for
$Z \leftarrow X^\top Y$
Compute QR decomposition: $Z = QR$, keep $Q \in \mathbf{R}^{D \times l}$ orthonormal
Truncate $Q \leftarrow Q[:, 1 : q]$
$Y \leftarrow XQ$
return Y, Q

E Experiments Additional Figures

E.1 Synthetic Results

E.1.1 Sparse Regime Performance Results

Table 6: Reconstruction error (sparse regime) across priors and sample sizes $N = D$.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	0.0254	0.0342	0.0232	0.0295	0.0237	0.0286
Student-t	0.0254	0.0342	0.0232	0.0295	0.0237	0.0286
Laplace	0.0254	0.0342	0.0232	0.0295	0.0237	0.0286
Log-quadratic	0.0254	0.0342	0.0232	0.0295	0.0237	0.0286
Spike-and-Slab	0.0254	0.0340	0.0232	0.0294	0.0237	0.0286

Table 7: Runtime (sparse regime) across priors and sample sizes $N = D$, comparing projected vs non-projected.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	1.52	4.67	1.90	21.8	2.56	124
Student-t	1.55	4.76	1.94	22.1	2.61	124
Laplace	1.50	4.67	1.88	21.8	2.53	124
Log-quadratic	1.53	4.71	1.91	22.1	2.54	125
Spike-and-Slab	1.79	5.07	2.14	22.6	2.88	130

Table 8: Factor recovery (sparse regime) across priors and sample sizes $N = D$, comparing projected vs non-projected.

Prior	$N = 1000$		$N = 2000$		$N = 4000$	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	0.173	0.196	0.197	0.199	0.185	0.204
Student-t	0.173	0.196	0.197	0.199	0.185	0.204
Laplace	0.173	0.195	0.198	0.198	0.185	0.204
Log-quadratic	0.173	0.196	0.197	0.199	0.185	0.204
Spike-and-Slab	0.178	0.195	0.196	0.201	0.181	0.192

E.2 Olivetti Faces

Table 9: Average final ELBO value comparing projected vs non-projected BB-VI PCA across priors and latent dimensions K.

Prior	K=10		K=20		K=30	
	Projected	Non-Projected	Projected	Non-Projected	Projected	Non-Projected
Gaussian	-19657.06	-702511.98	-31502.83	-642086.99	-42533.59	-644319.76
Student-t	-19653.65	-703331.65	-31511.50	-643521.96	-42549.78	-646477.56
Laplace	-19685.03	-712885.91	-31611.68	-661179.18	-42711.55	-672645.97
Log-quadratic	-19657.34	-704783.31	-31528.89	-646057.33	-42572.52	-650285.94
Spike-and-Slab	-19756.65	-726424.13	-31453.81	-645747.86	-42636.06	-610385.34

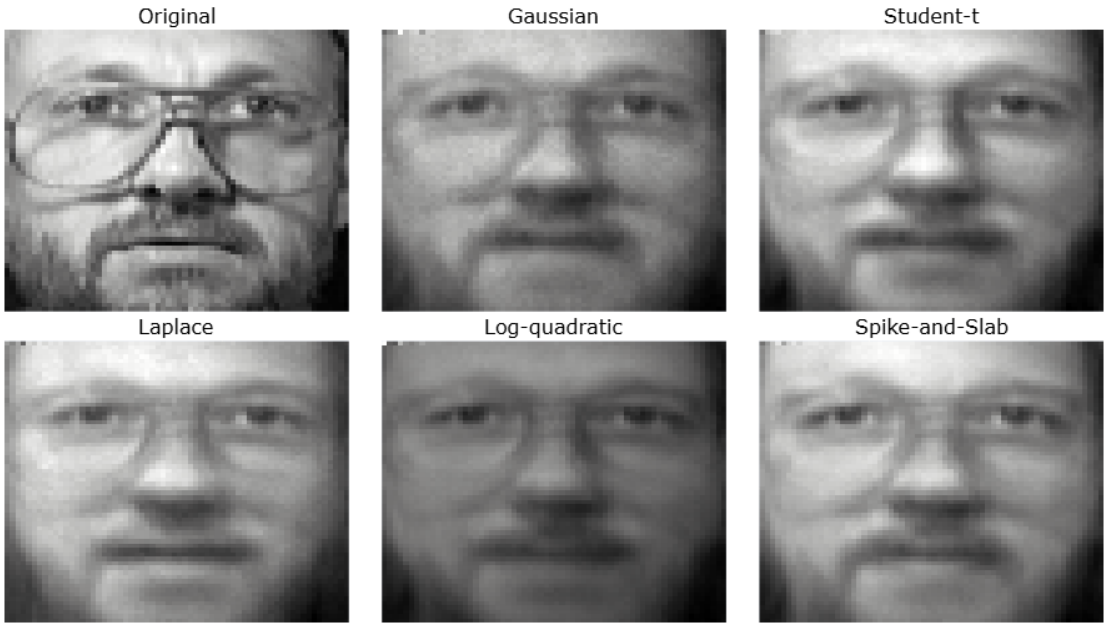

Figure 2: Sample Olivetti faces reconstruction (K=20) using Bayesian PCA with randomized projections across priors.

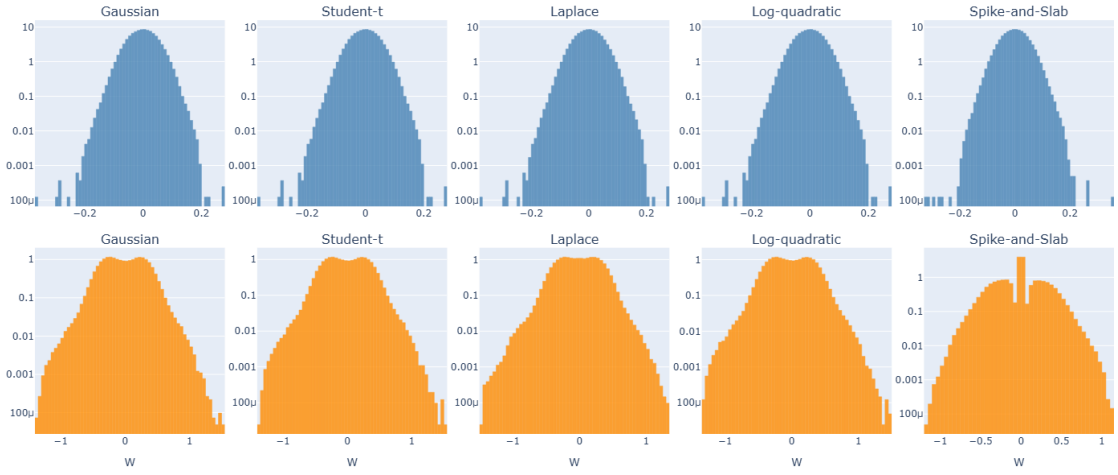

Figure 3: Posterior distributions of loading matrix W on Olivetti Faces ($K = 20$), comparing projected (top) and unprojected (bottom) inference across priors.

References

[1] Christopher Bishop. Bayesian pca. In *Advances in Neural Information Processing Systems*, volume 11. MIT Press, 1998.

[2] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.

[3] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. In *SIAM review*, volume 53, pages 217–288. SIAM, 2011.

[4] William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(121-129):189, 1984.

[5] Rajesh Ranganath, Sean Gerrish, and David M Blei. Black box variational inference. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2014.

[6] Farzin Samaria and Andy Harter. Parameterisation of a stochastic model for human face recognition. *Proc. Second IEEE Int. Workshop on Automatic Face and Gesture Recognition*, pages 138–142, 1994.

[7] Tamas Sarlos. Improved approximation algorithms for large matrices via random projections. *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 143–152, 2006.