AMSI VACATION RESEARCH SCHOLARSHIPS 2020–21

Get a Thirst for Research this Summer

Variational Inference for Bayesian Nonnegative Matrix Factorisation

Gyu Hwan Park
Supervised by Heejung Shim
University of Melbourne

Vacation Research Scholarships are funded jointly by

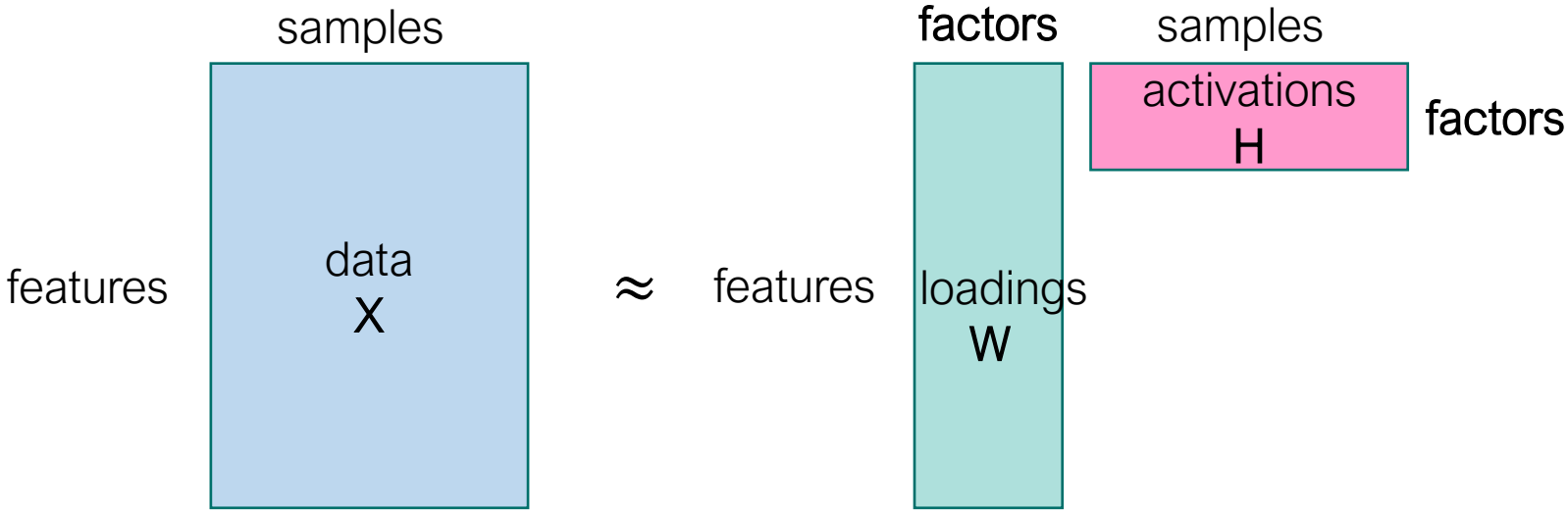
AMSI Australian Mathematical Sciences Institute

Australian Government
Department of Education, Skills and Employment

AMSI CONNECT

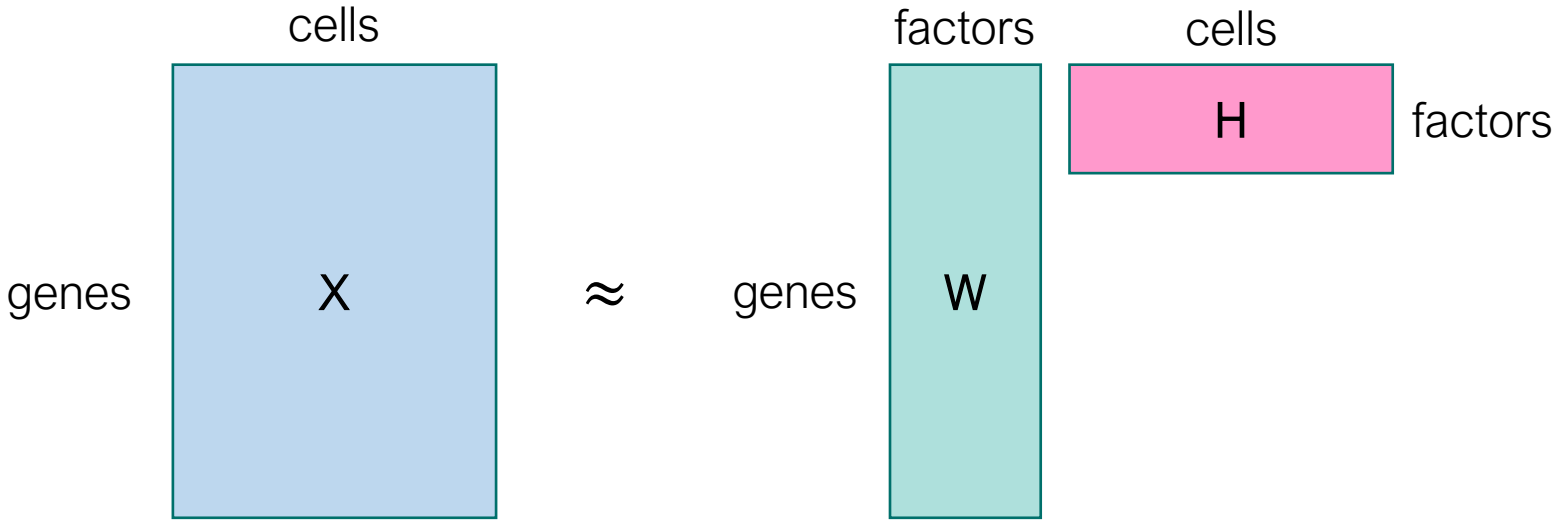
Nonnegative Matrix Factorisation (NMF)

- Decompose data matrix (**X**) as a product of two nonnegative matrices of factor loadings (**W**) and activations (**H**)
- Nonnegativity constraint for interpretation



Example: single-cell RNA-seq data

- Identify underlying biological processes (factors) in gene expression data
- How can we form meaningful interpretation?

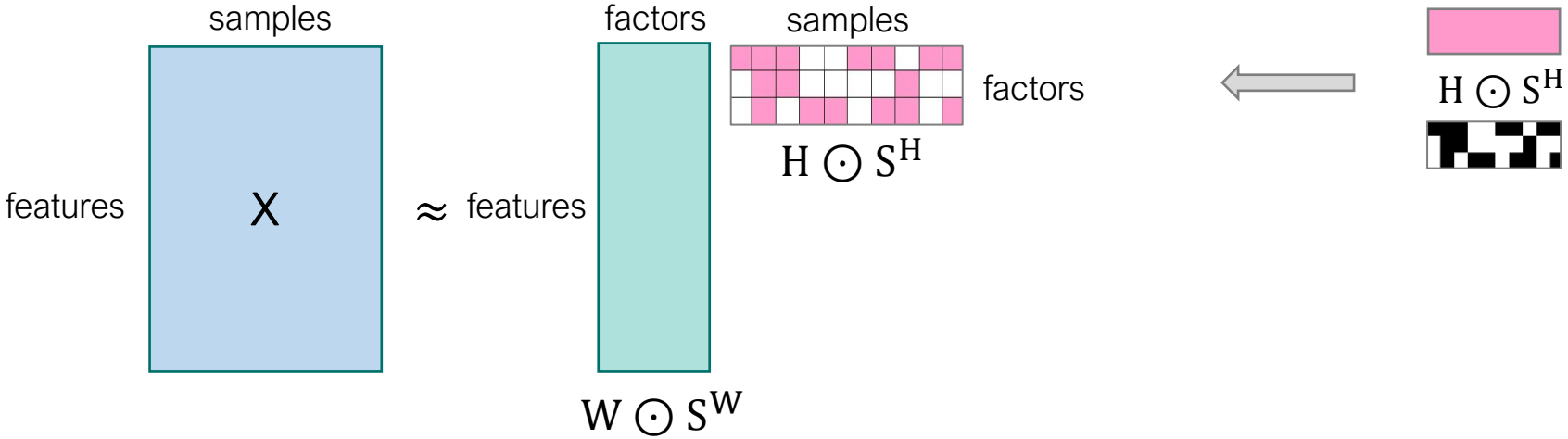


Sparse NMF (S-NMF)

- Binary mask S^H imposes sparsity on H

$$X \approx W(H \odot S^H)$$

- Idea: factors may only be associated with a fraction of cells
- Liang et al. (2013)

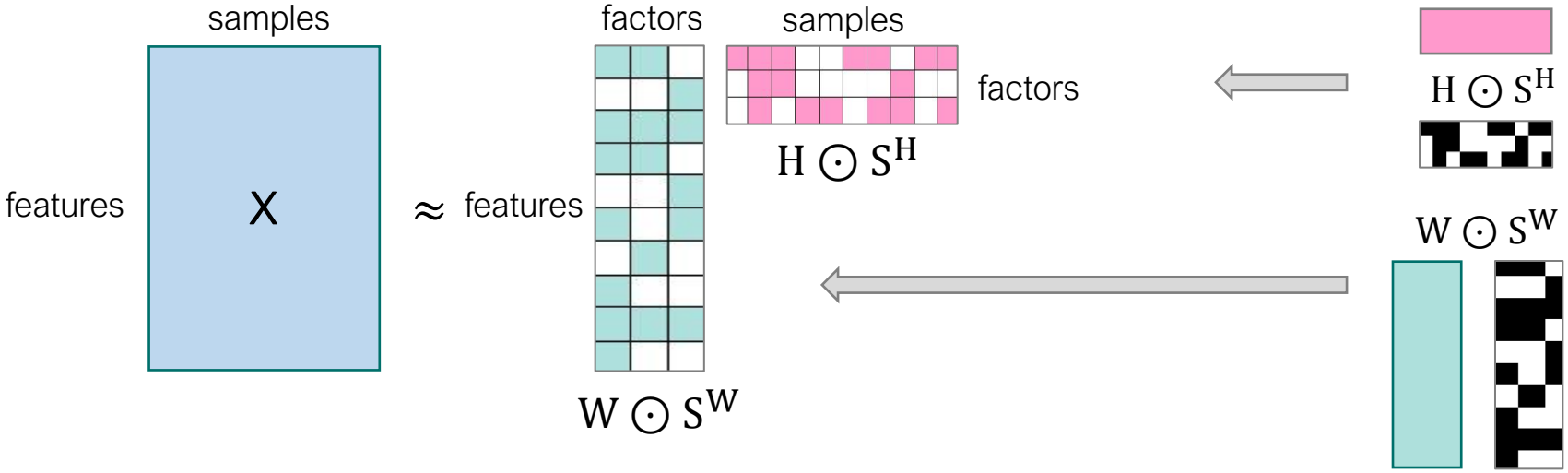


Doubly Sparse NMF (DS-NMF)

- Additional binary mask $\mathbf{S^W}$ imposes sparsity on $\mathbf{W}$

$$X \approx (W \odot S^W)\,(H \odot S^H)$$

- Learn from data which factors affect which subset of features and samples



Bayesian inference

Joint probability likelihood prior

posterior distribution $p(\theta \mid x) = \dfrac{p(x,\theta)}{p(x)} = \dfrac{p(x \mid \theta)\, p(\theta)}{\int p(x \mid \theta)\, p(\theta)\, d\theta}$

marginal probability

- Posterior is typically intractable to compute
- Resort to approximate methods like Markov Chain Monte Carlo (MCMC) or **Variational Inference (VI)**

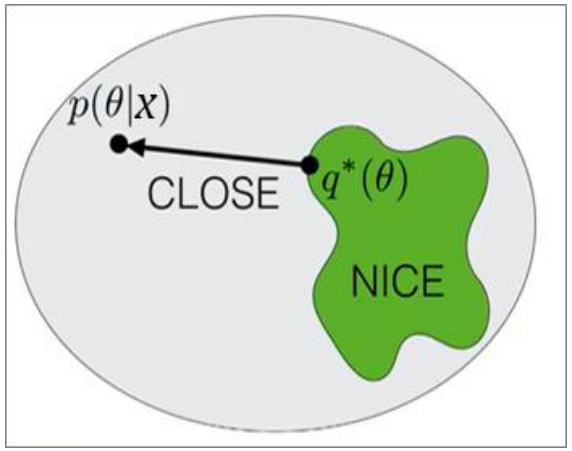
Variational Inference (VI)

- Approximate the posterior distribution with '**closest**' variational distribution $q^*(\theta)$ from a '**nice**' family of distributions

$$q^*(\theta) = \operatorname{argmin}_{q \in Q} \boldsymbol{KL}\big(\boldsymbol{q}(\boldsymbol{\theta}) \,|\, \boldsymbol{p}(\boldsymbol{\theta}|\boldsymbol{x})\big)$$

- Equivalent to maximising the *Evidence Lower Bound (ELBO)*

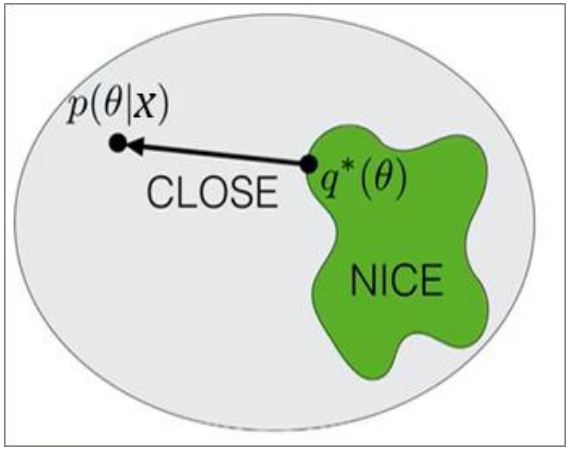
$$\mathrm{ELBO}(q) = \mathbb{E}_q\left[\log q(\theta) - \log p(x, \theta)\right]$$



Source: Broderick

Variational Inference (VI)

- Common choice for family **Q**: mean-field variational family
  - Assumption: independence of parameters in variational distribution
- Enables tractable, often closed-form iterative optimisation

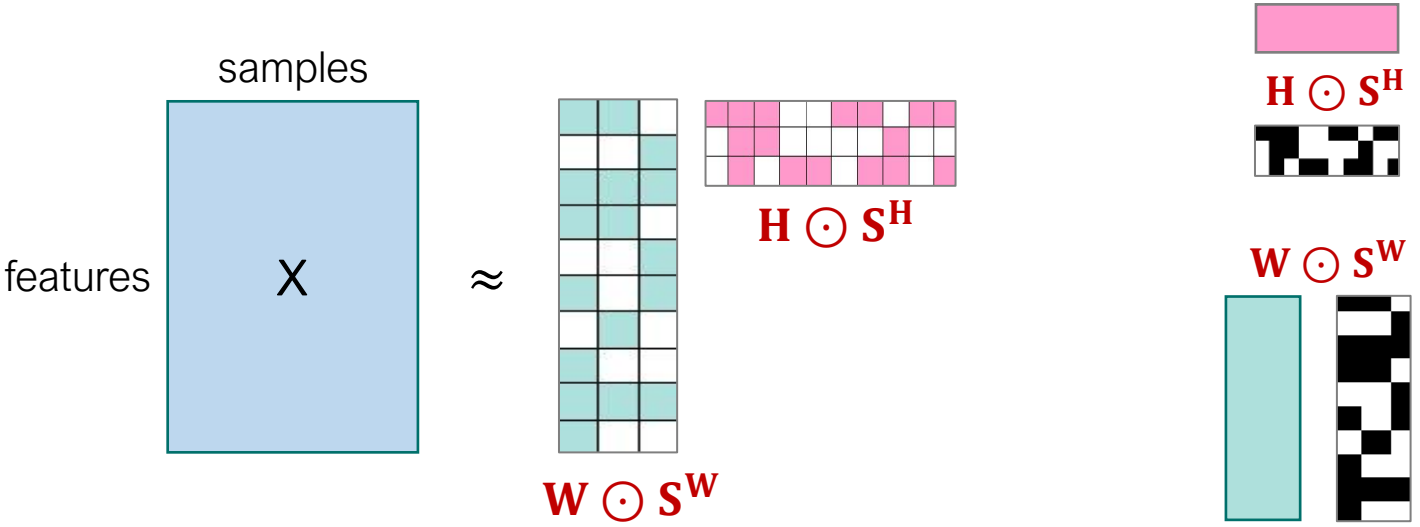


Source: Broderick

DS-NMF with VI

$$X \approx (W \odot S^W)\,(H \odot S^H)$$

- Independence assumed between W and S^W, H and S^H

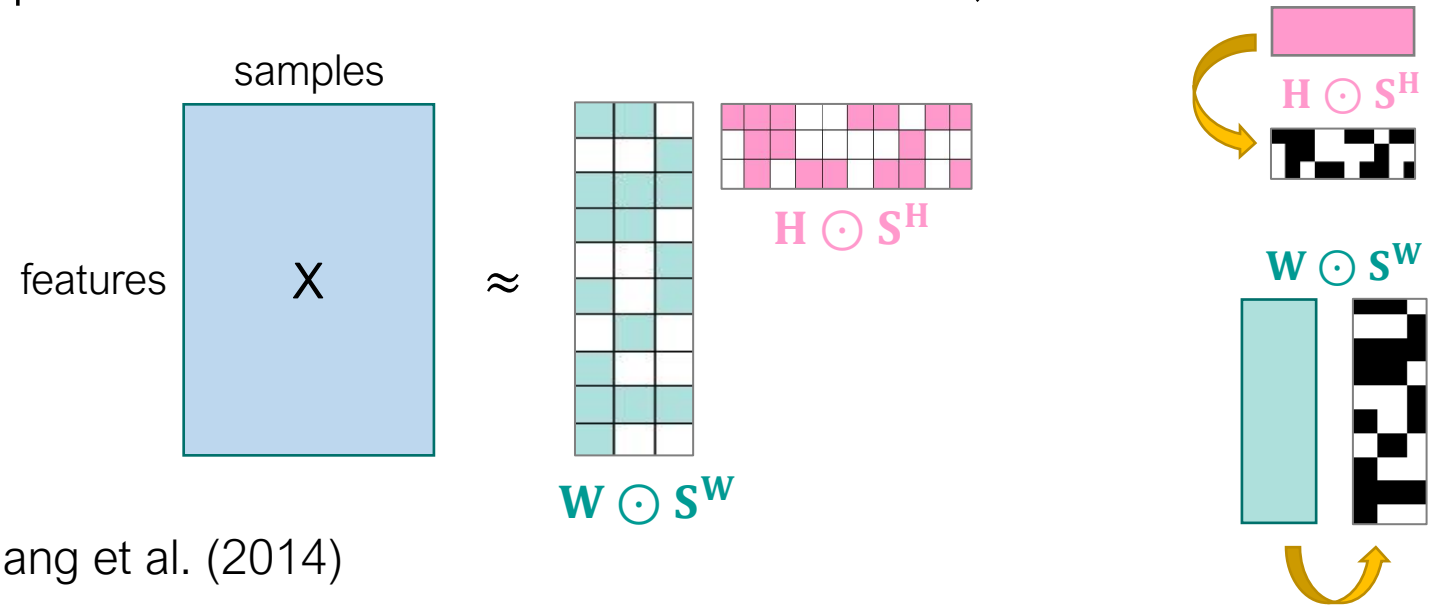


DS-NMF with Structured Stochastic VI (DSSVI)

- Inference: *Structured Stochastic Variational Inference*

$$X \approx (W \odot S^W)\,(H \odot S^H)$$

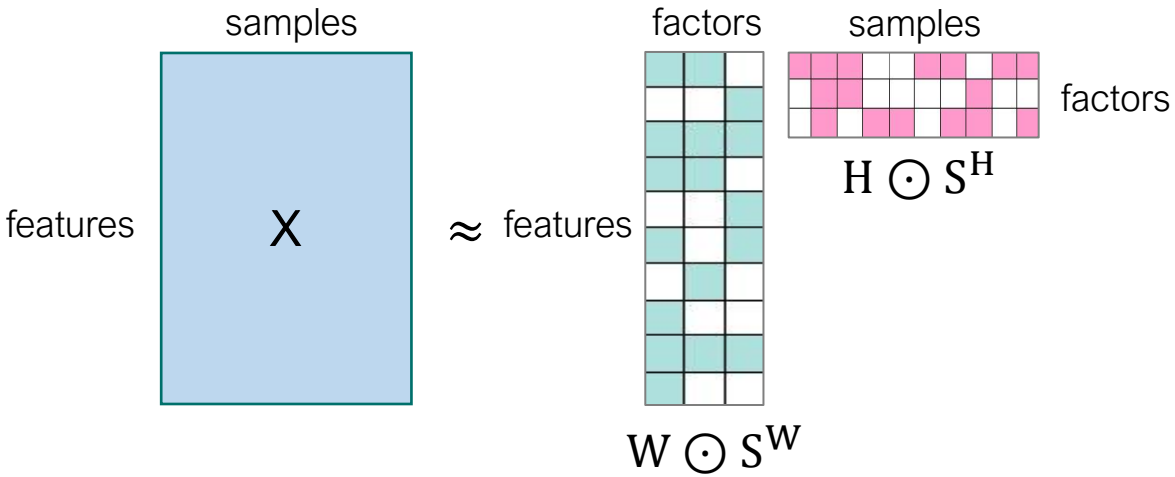
- Dependencies restored between W and S^W, H and S^H



- Liang et al. (2014)

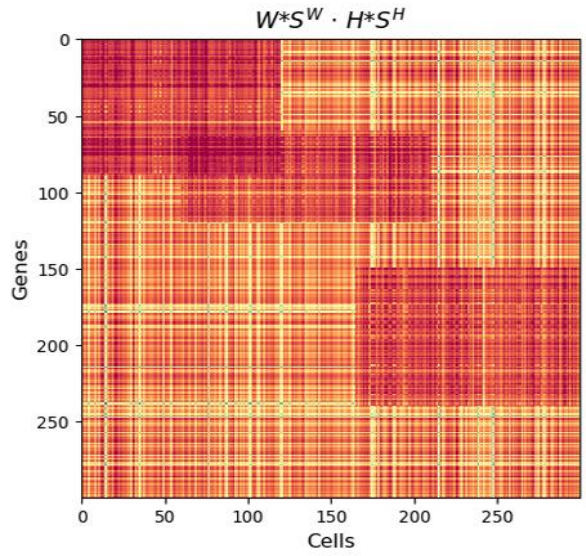
Simulation study

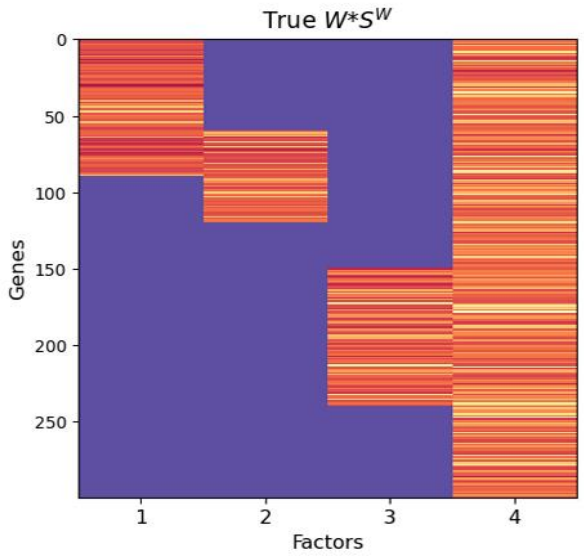
- Assess performance of DSSVI on estimating sparse S^W
- Compare **DSSVI vs SSVI** on estimating:
  - Factor loadings $W \odot S^W$
  - Sparse binary mask S^H
  - Activations $H \odot S^H$

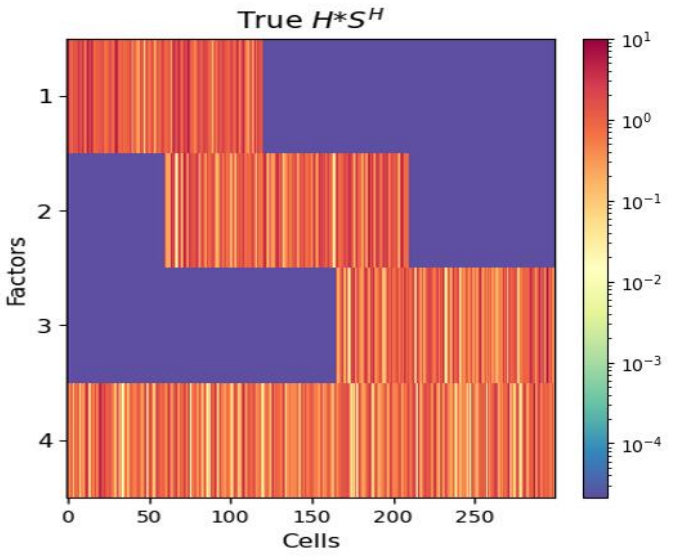


Simulated dataset

- 300 genes (features)
- 300 cells (samples)
- 4 factors



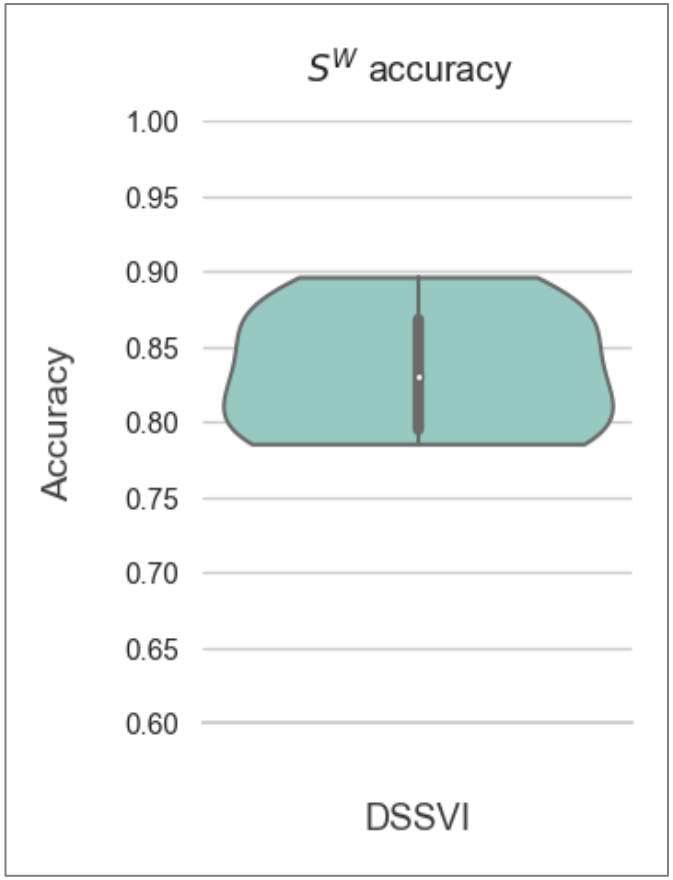




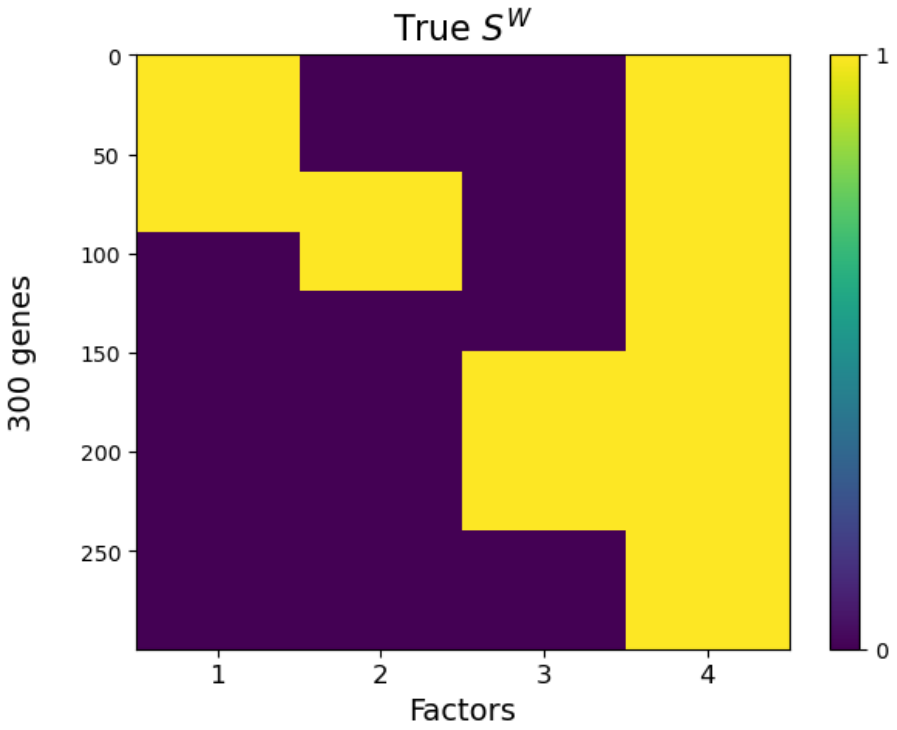
$E(X) = (W \odot S^W)(H \odot S^H)$

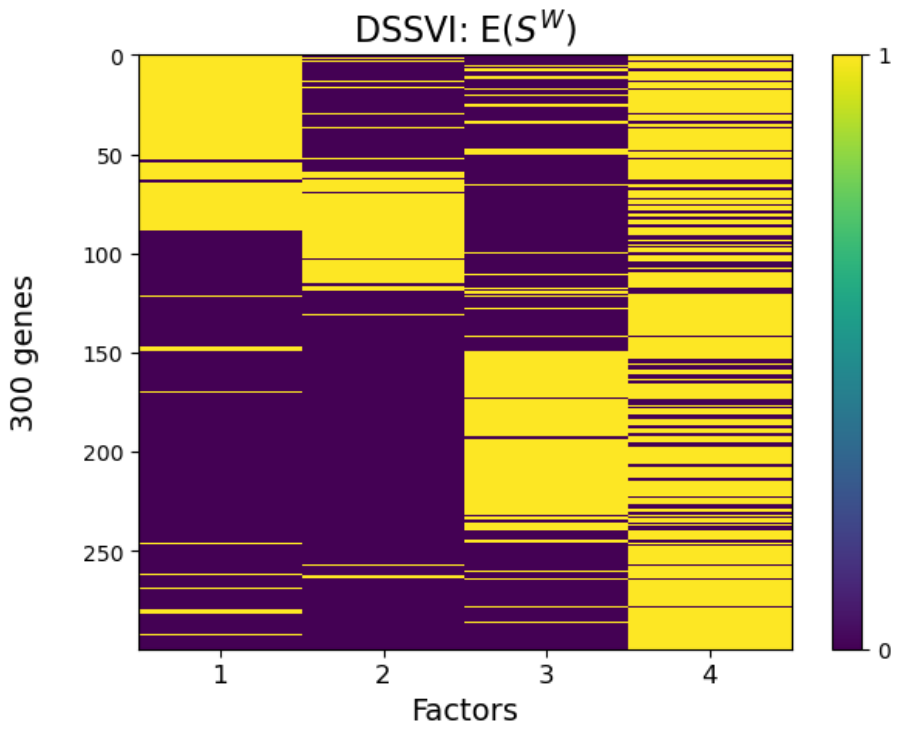
DSSVI: Capturing sparsity in S^W

- Use posterior mean for evaluation
- Accuracy of binary mask S^W: proportion of correctly inferred 0's and 1's



DSSVI: Estimated S^W



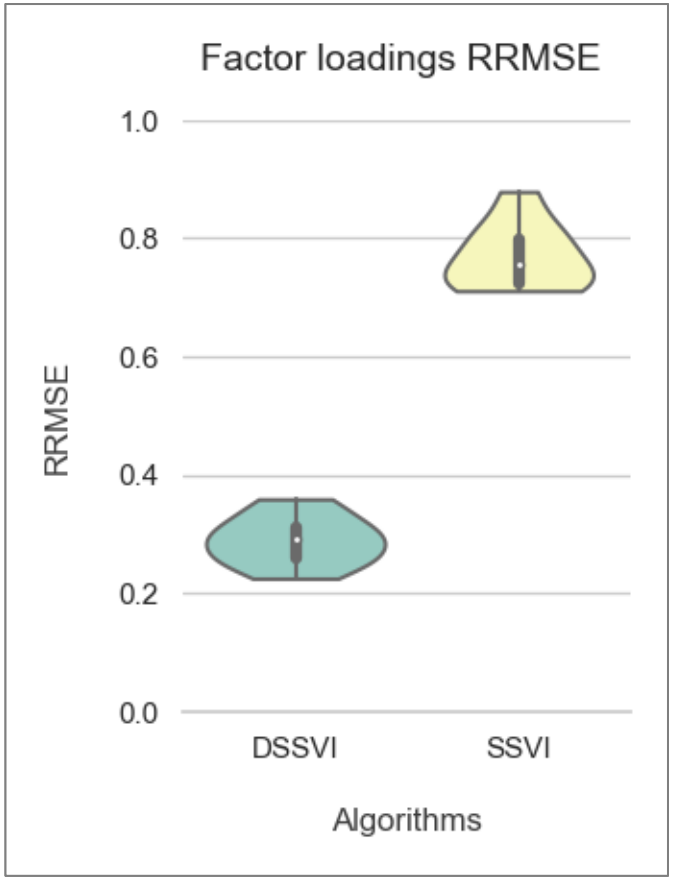


Estimation of sparse factor loadings

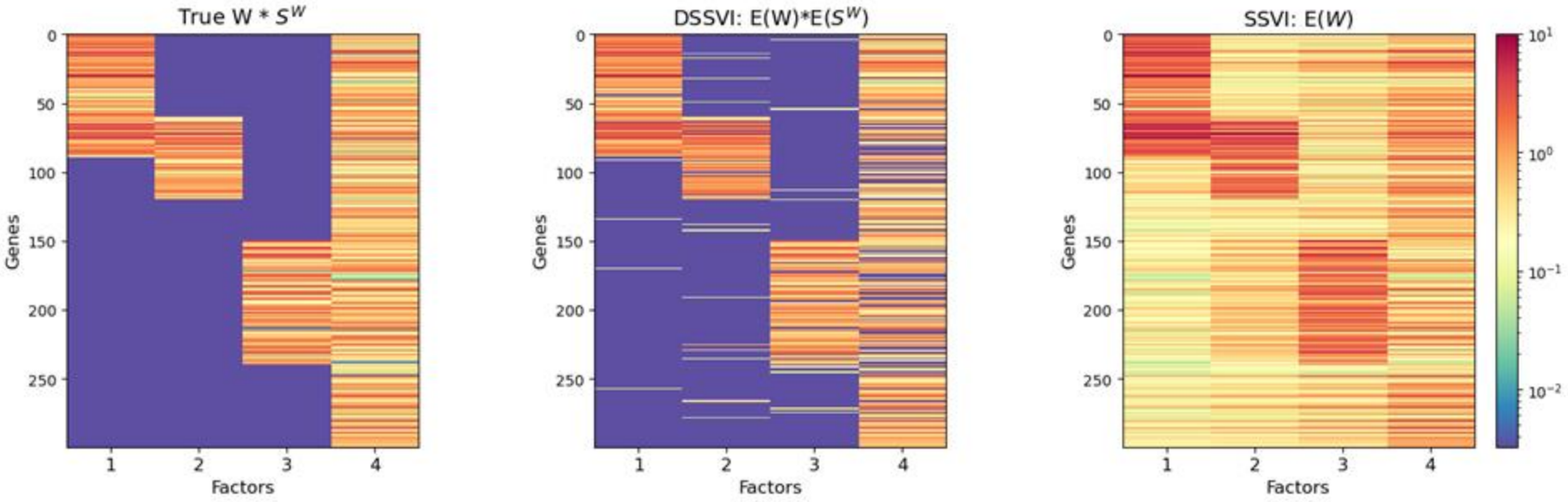
- *Relative Root Mean Squared Error (RRMSE):*

$$\mathrm{RRMSE}(\hat{A}, A) = \sqrt{\frac{\sum (A_{f,k} - \hat{A}_{f,k})^2}{\sum A_{f,k}^2}}$$

- DSSVI can estimate $W \odot S^W$ better



Estimation of sparse factor loadings



- SSVI fails to capture the sparse structure of factor loadings
- DSSVI can identify which fraction of genes are affected by which factors

Discussion

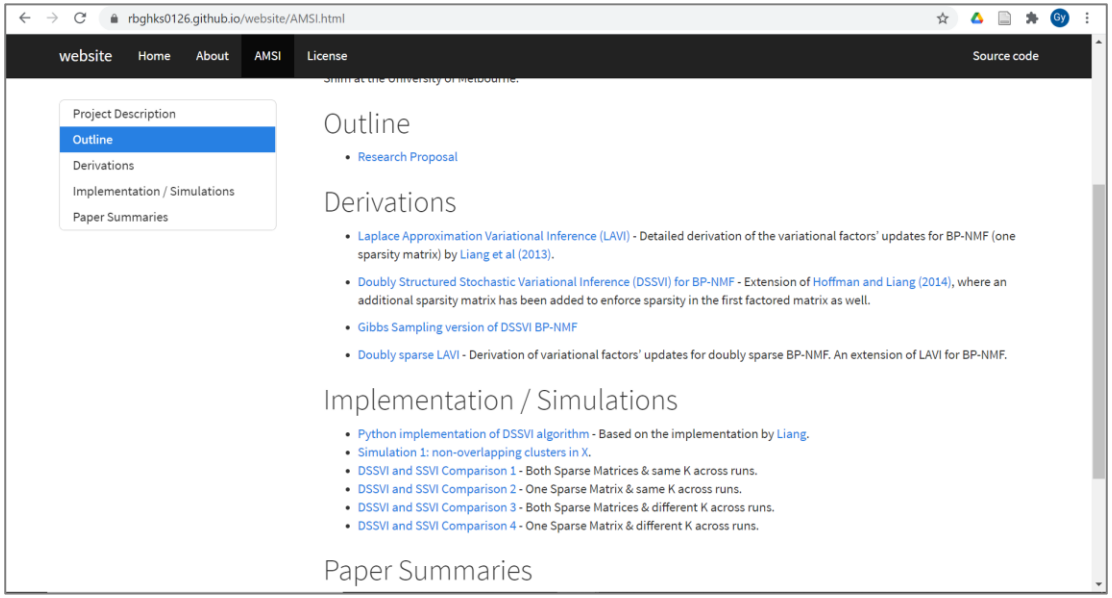
- DSSVI can capture sparsity in S^H as good as SSVI
- DSSVI performs better in estimating mean of observations
- DSSVI can flexibly learn whether or not sparsity is present in true data

Next steps

- Apply to real data
  - Single cell RNA-seq data
- Implement log-predictive likelihood metric for evaluation

Conclusion

- DSSVI enhances performance and interpretation
  - Can capture sparsity in factor loadings well
  - Can capture sparsity in activations as good as SSVI
- More details: https://rbghks0126.github.io/website/AMSI.html



References

D. Blei, A. Kucukelbir, and J. McAuliffe (2017). Variational Inference: A Review for Statisticians. Journal of the American Statistical Association.

D. Liang, M. D. Hoffman and D. Ellis (2013). Beta Process Sparse Nonnegative Matrix Factorization for Music. ISMIR.

D. Liang and M. D. Hoffman (2014). Beta process non-negative matrix factorization with stochastic structured mean-field variational inference. [online] Available: https://arxiv.org/abs/1411.1804.

Tamara Broderick: Variational Bayes and Beyond: Bayesian Inference for Big Data (ICML 2018 tutorial) (2018). YouTube video, added by Steven Van Vaerenbergh [online]. Available: https://www.youtube.com/watch?v=Moo4-KR5qNg&t=1502s&ab_channel=StevenVanVaerenbergh