Fast Bayesian Nonnegative Matrix Factorisation and Tri-Factorisation

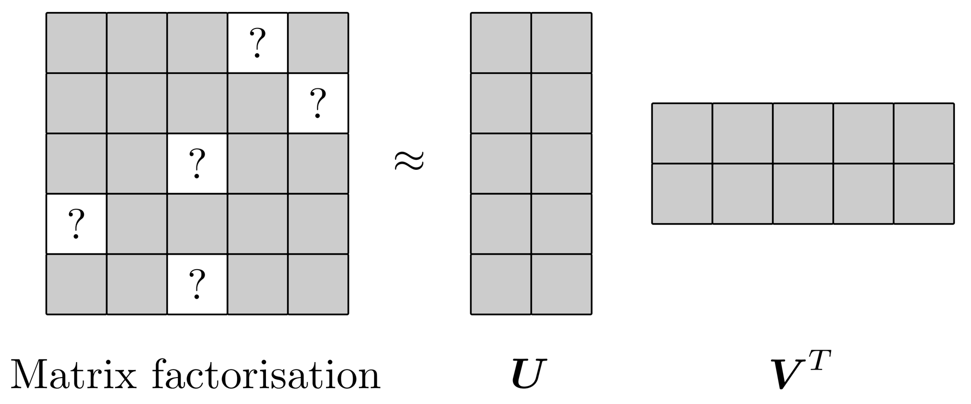
Thomas Brouwer[†], Jes Frellsen, Pietro Liò

University of Cambridge, [†]tab43@cam.ac.uk

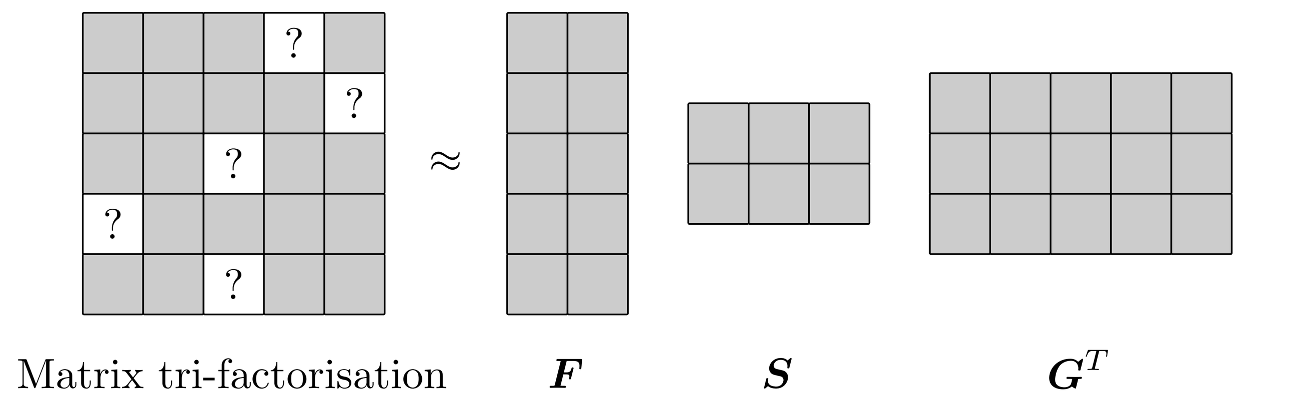
Nonnegative matrix factorisation and tri-factorisation

Nonnegative matrix factorisation (NMF) and tri-factorisation (NMTF) methods decompose a given matrix $\boldsymbol{R}$ into two or three smaller matrices so that $\boldsymbol{R} \approx \boldsymbol{U}\boldsymbol{V}^T$ or $\boldsymbol{R} \approx \boldsymbol{F}\boldsymbol{S}\boldsymbol{G}^T$, respectively. Schmidt, Winther and Hansen (2009) introduced a Bayesian version of nonnegative matrix factorisation (left), which we extend to matrix tri-factorisation (right).

$$R_{ij} \sim \mathcal{N}(R_{ij}|\boldsymbol{U}_i \cdot \boldsymbol{V}_j, \tau^{-1}) \qquad \tau \sim \mathcal{G}(\tau|\alpha,\beta)$$
$$U_{ik} \sim \mathcal{E}(U_{ik}|\lambda_U) \qquad V_{jk} \sim \mathcal{E}(V_{jk}|\lambda_V)$$



$$R_{ij} \sim \mathcal{N}(R_{ij}|\boldsymbol{F}_i \cdot \boldsymbol{S} \cdot \boldsymbol{G}_j, \tau^{-1}) \qquad \tau \sim \mathcal{G}(\tau|\alpha,\beta)$$
$$F_{ik} \sim \mathcal{E}(F_{ik}|\lambda_F) \qquad S_{kl} \sim \mathcal{E}(S_{kl}|\lambda_S) \qquad G_{jl} \sim \mathcal{E}(G_{jl}|\lambda_G)$$



Slow inference – Gibbs sampling

Schmidt et al. introduced a Gibbs sampling algorithm for inference, to approximate the posterior distribution over $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{F}, \boldsymbol{S}, \boldsymbol{G}$. We sample new values randomly for each entry in turn of the posteriors given below (for NMF), to converge to the true posterior. The parameter values can be derived using Bayes' theorem.

$$p(\tau|\boldsymbol{U},\boldsymbol{V},D) = \mathcal{G}(\tau|\alpha^*,\beta^*)$$
$$p(U_{ik}|\tau,\boldsymbol{U}_{-ik},\boldsymbol{V},D) = \mathcal{TN}(U_{ik}|\mu_{ik}^U,\tau_{ik}^U)$$
$$p(V_{jk}|\tau,\boldsymbol{U},\boldsymbol{V}_{-jk},D) = \mathcal{TN}(V_{jk}|\mu_{jk}^V,\tau_{jk}^V)$$

$\mathcal{TN}$ is a truncated normal (Gaussian with zero density below $x = 0$). If instead of random draws we use the mode, we get a MAP solution (iterated conditional modes, ICM).

Fast inference – Variational Bayes

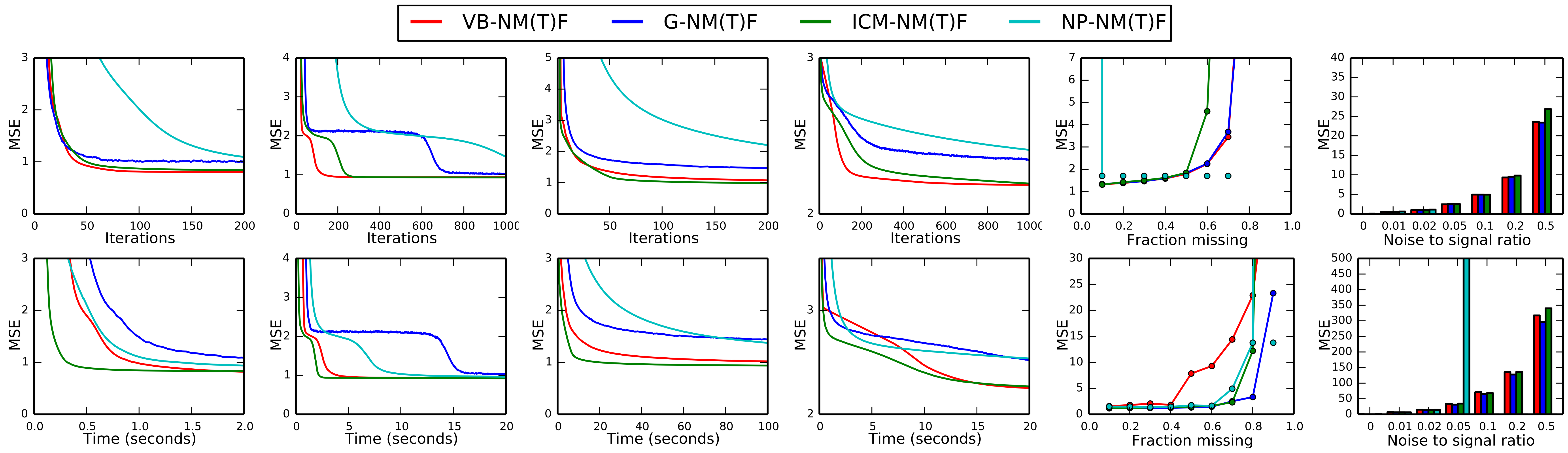
Variational Bayesian inference (VB) is an alternative to Gibbs sampling, where we approximate the true posterior $p(\boldsymbol{\theta}|D)$ with an approximation $q(\boldsymbol{\theta})$ that is easier to compute. We make the mean-field assumption, so all variables in our approximation are independent. We choose the posteriors as follows:

$$q(\tau) = \mathcal{G}(\tau|\alpha^*,\beta^*)$$
$$q(U_{ik}) = \mathcal{TN}(U_{ik}|\mu_{ik}^U,\tau_{ik}^U)$$
$$q(V_{jk}) = \mathcal{TN}(V_{jk}|\mu_{jk}^V,\tau_{jk}^V)$$

VB does not rely on random draws, instead solving an optimisation problem, and has two advantages: it can convergence much faster, and does not require additional draws to approximate the posterior.

Experiments

Methods NMF and NMTF: Gibbs, ICM, and VB; non-probabilistic NMF (Lee and Seung 2001); non-probabilistic NMTF (Yoo and Choi 2009). **Experiments**: • Convergence speed on simulated data, and a drug sensitivity dataset (GDSC). • Missing values predictions test with varying fractions of missing entries and noise levels.



Columns 1-4: Convergence of algorithms on toy (NMF: 1, NMTF: 2) and GDSC drug sensitivity (NMF: 3, NMTF: 4) data, measuring training data fit (MSE) across iterations (top) and time (bottom). **Column 5-6**: Missing values prediction performances (5) and noise test performances (6), measuring average predictive performance on test set (MSE) for different fractions of unknown values and noise-to-signal ratios. Top: NMF, bottom: NMTF.

Conclusion

- We have introduced a faster inference algorithm for Bayesian nonnegative matrix factorisation, using variational Bayesian inference, and shown that it offers superior rates of inference. It is competitive with a MAP method, yet gives a full posterior approximation.
- We also introduced a Bayesian version of nonnegative matrix tri-factorisation, where inference is even harder. The fast variational Bayesian approach opens up the application of BNMTF to bigger datasets and future extensions.

References

M. N. Schmidt, O. Winther, and L. K. Hansen. **Bayesian non-negative matrix factorization**. In *International Conference on Independent Component Analysis and Signal Separation*, 2009.

J. Yoo and S. Choi. **Probabilistic matrix tri-factorization**. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009.

D. D. Lee and H. S. Seung. **Algorithms for Non-negative Matrix Factorization**. *NIPS 2000, MIT Press.*