Comparison of variational Bayes and Gibbs sampling in reconstruction of missing values with probabilistic principal component analysis

Luis De Alba, Alexander Ilin, and Tapani Raiko
Adaptive Informatics Research Center
Aalto University

Collaborative filtering

- Consider a table of ratings of movies (m1,m2,...) given by different people (p1,p2,...)
- The task is to make personalized recommendations
- This can be seen as reconstructing missing ratings in the table given the observed ones

	p1	**p2**	**p3**	**p4**	**p5**	**p6**
m1	4	5	**?**	5	2	
m2	5	**?**		5		3
m3		1	3		5	**?**

Principal Component Analysis (PCA)

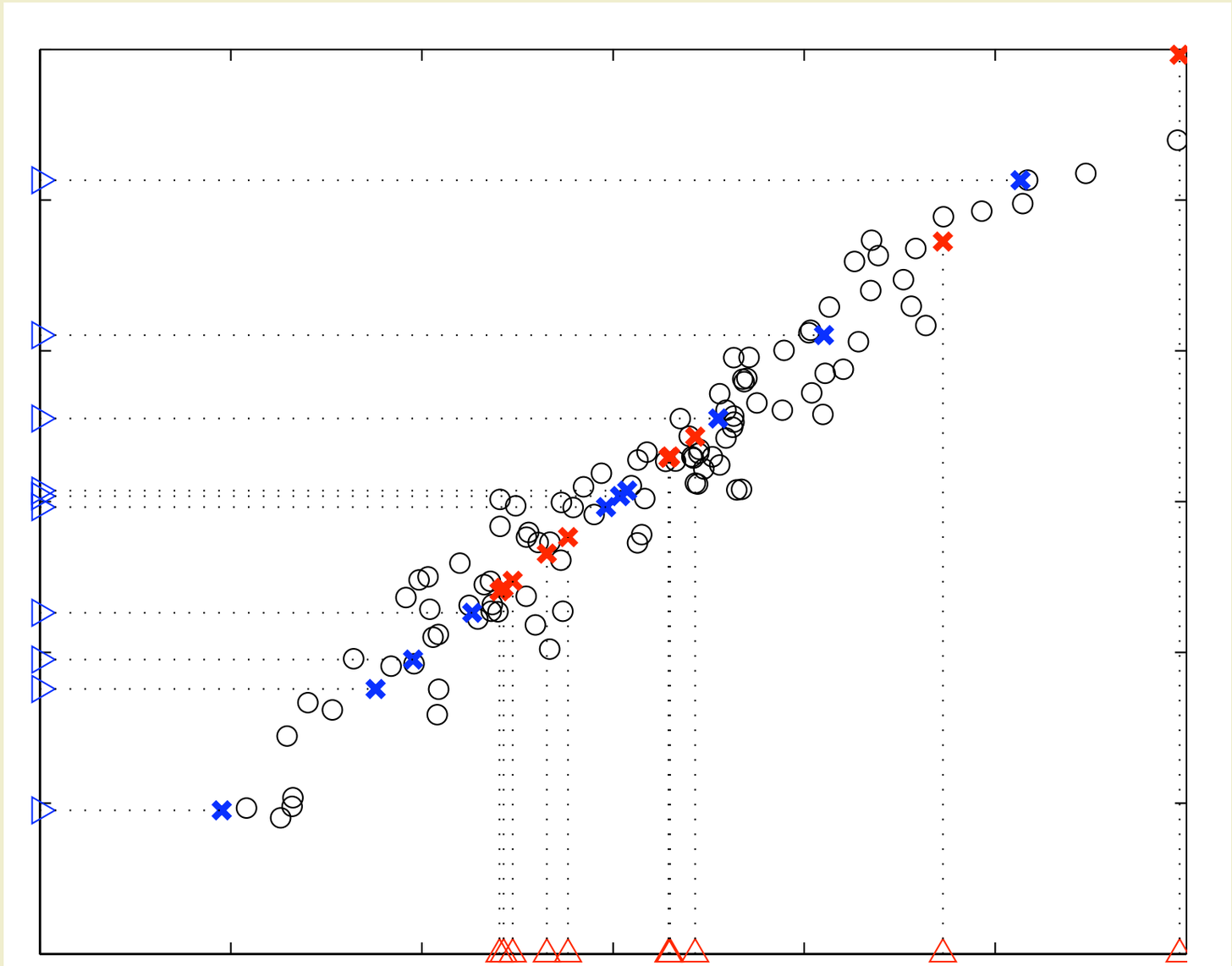
- Data Y consists of n d-dimensional vectors
- Matrix Y is decomposed in to a product of smaller matrices such that the square reconstruction error is minimized

$$\mathbf{Y} \approx \mathbf{W}\mathbf{X} + \mathbf{m}$$

Algorithms for PCA

- Eigenvalue decomposition (standard approach)
  - Compute the covariance matrix and its eigenvectors
- EM algorithm
  - Iterates between updates of W and X

PCA with missing values

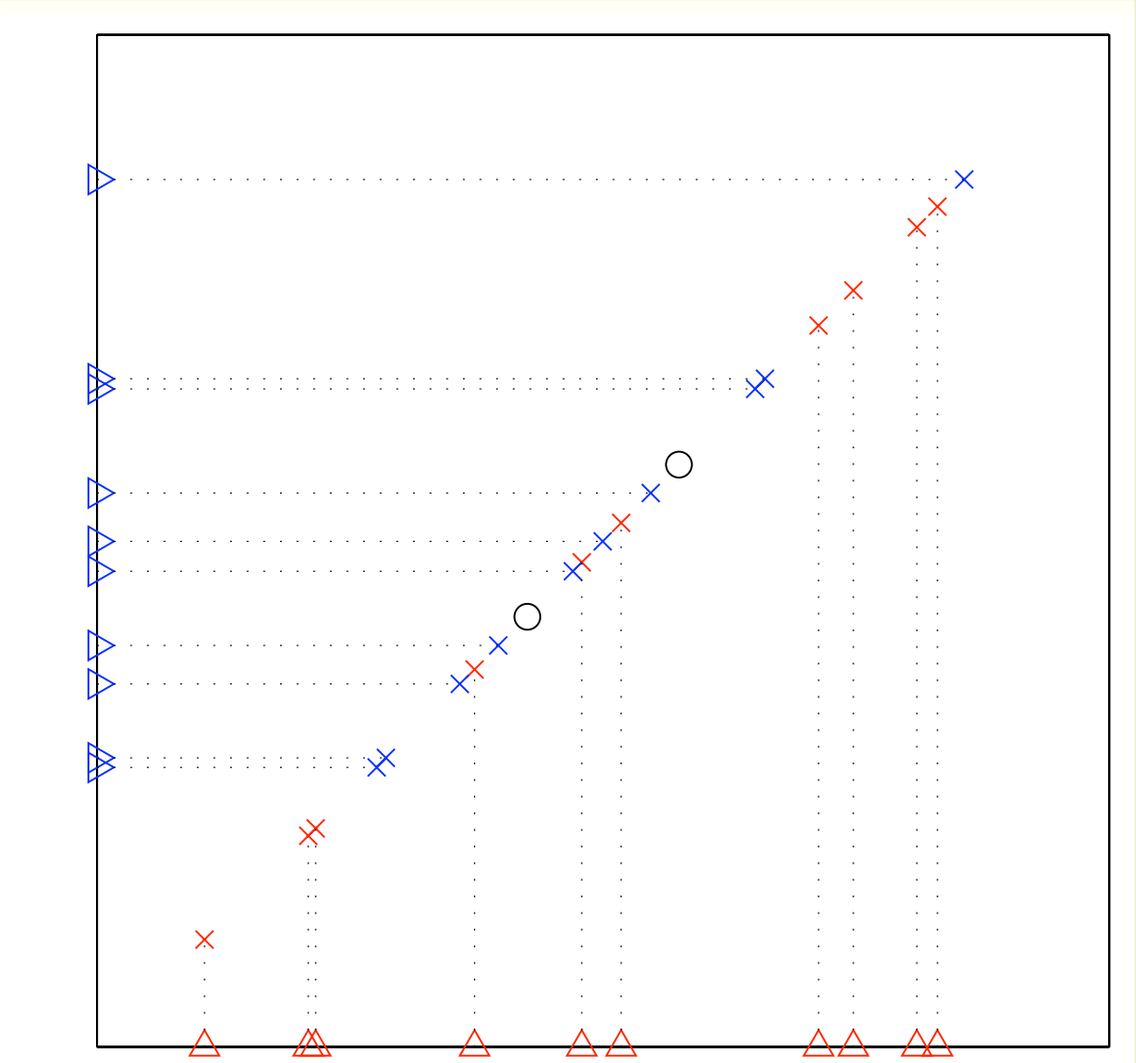


- Red and blue data points are reconstructed based on only one of the two dimensions

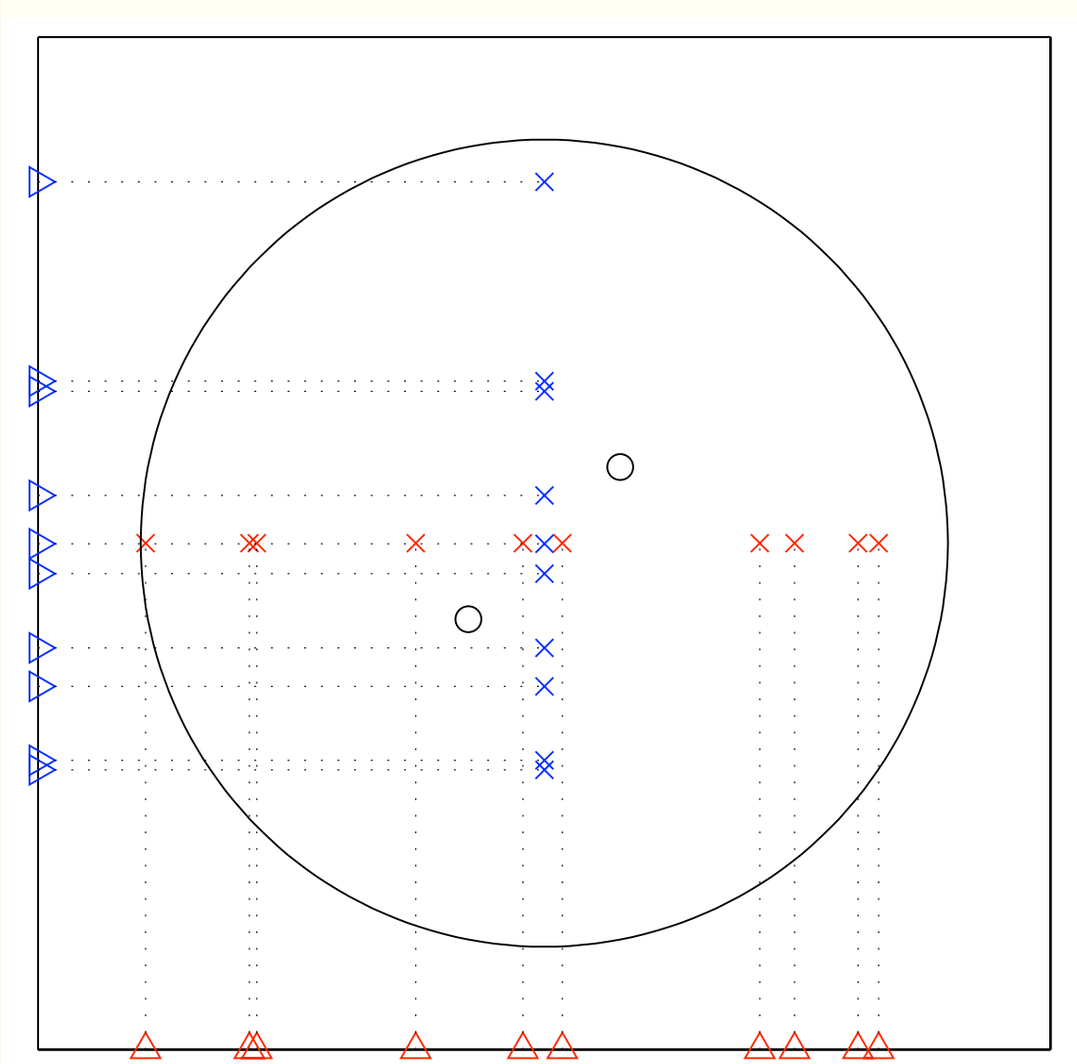
Adapting the algorithms for missing values

- Iterative imputation
  - Alternately 1) fill in missing values and
    2) solve normal PCA with the standard approach
- EM algorithm becomes a bit more involved
  - Can be extended, and was thus used here

Overfitting in case of sparse data

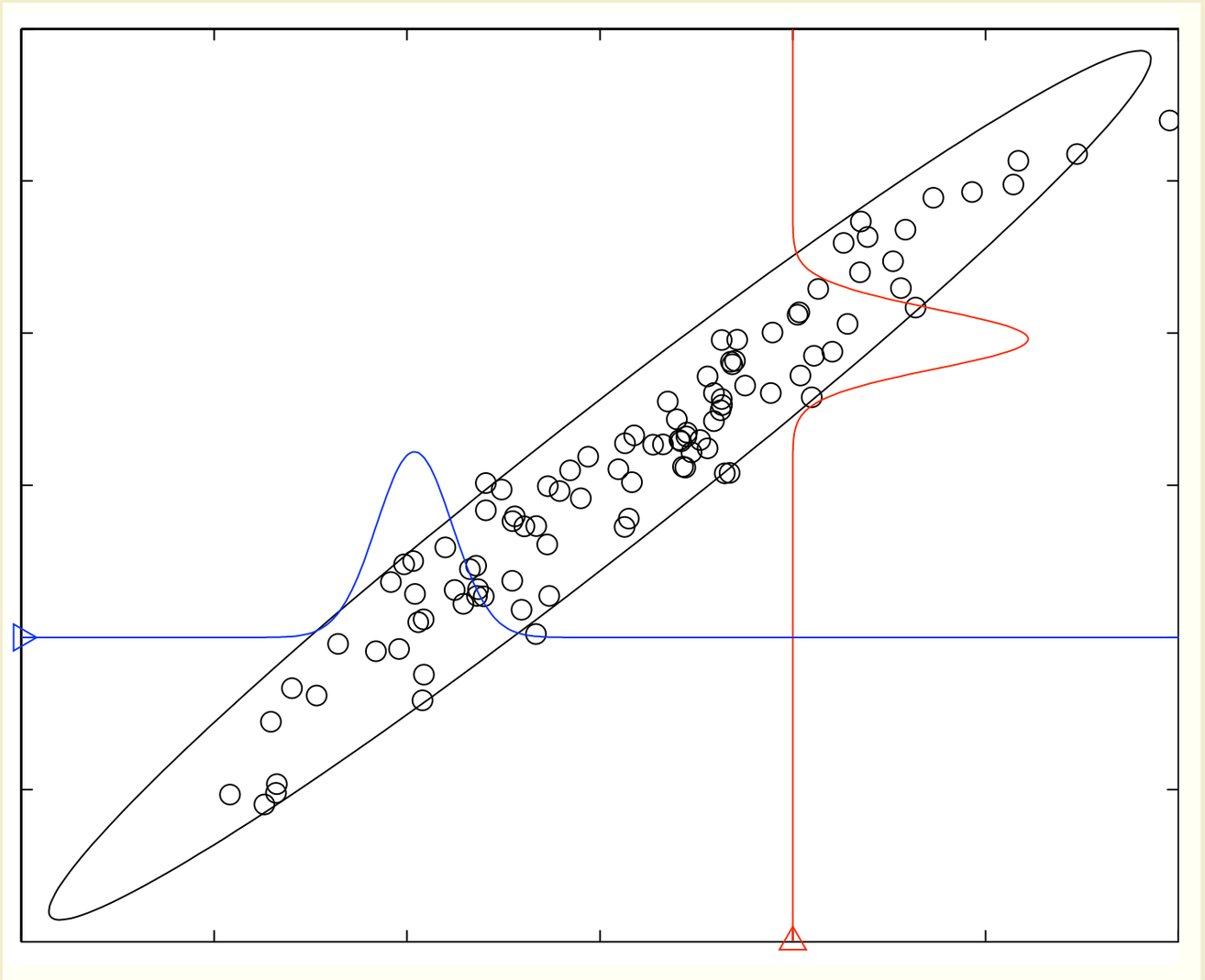


Overfitted solution



Regularized solution

Regularization against Overfitting



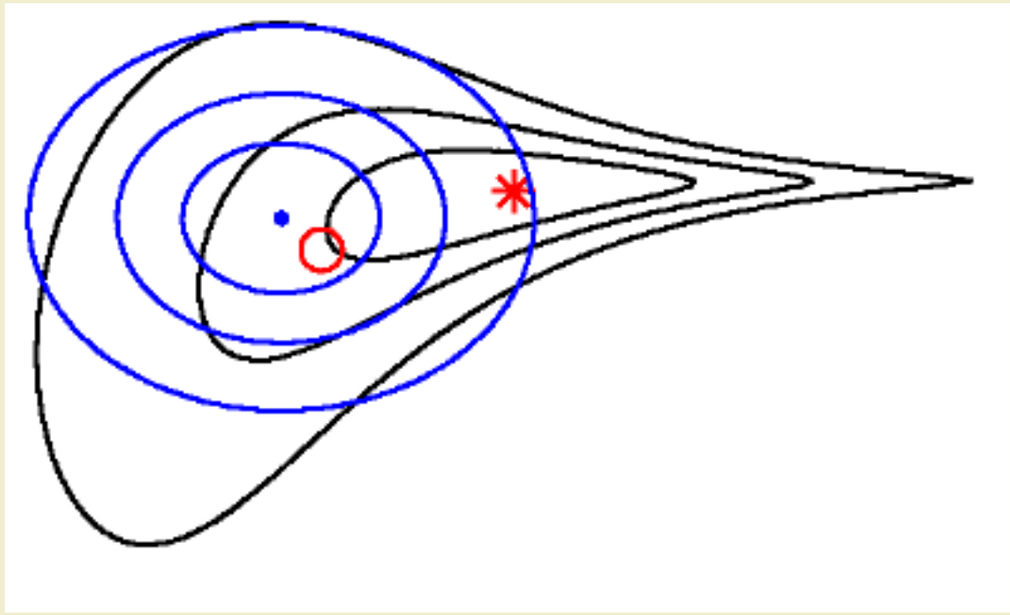
- Penalizing the use of large parameter values
- Estimating the distribution of unknown parameters

Modelling the uncertainty

$$P(\mathbf{Y}_{MIS}|\mathbf{Y}_{OBS}) = \int P(\mathbf{Y}_{MIS}|\mathbf{W},\mathbf{X},\mathbf{m})$$
$$P(\mathbf{W},\mathbf{X},\mathbf{m}|\mathbf{Y}_{OBS})\, d\mathbf{W}\, d\mathbf{X}\, d\mathbf{m}.$$

- The optimal reconstruction is an integral over the unknown variables of the model
- There is no analytical solution to the integral so approximative methods must be used
- Variational Bayes and Gibbs sampling are such methods

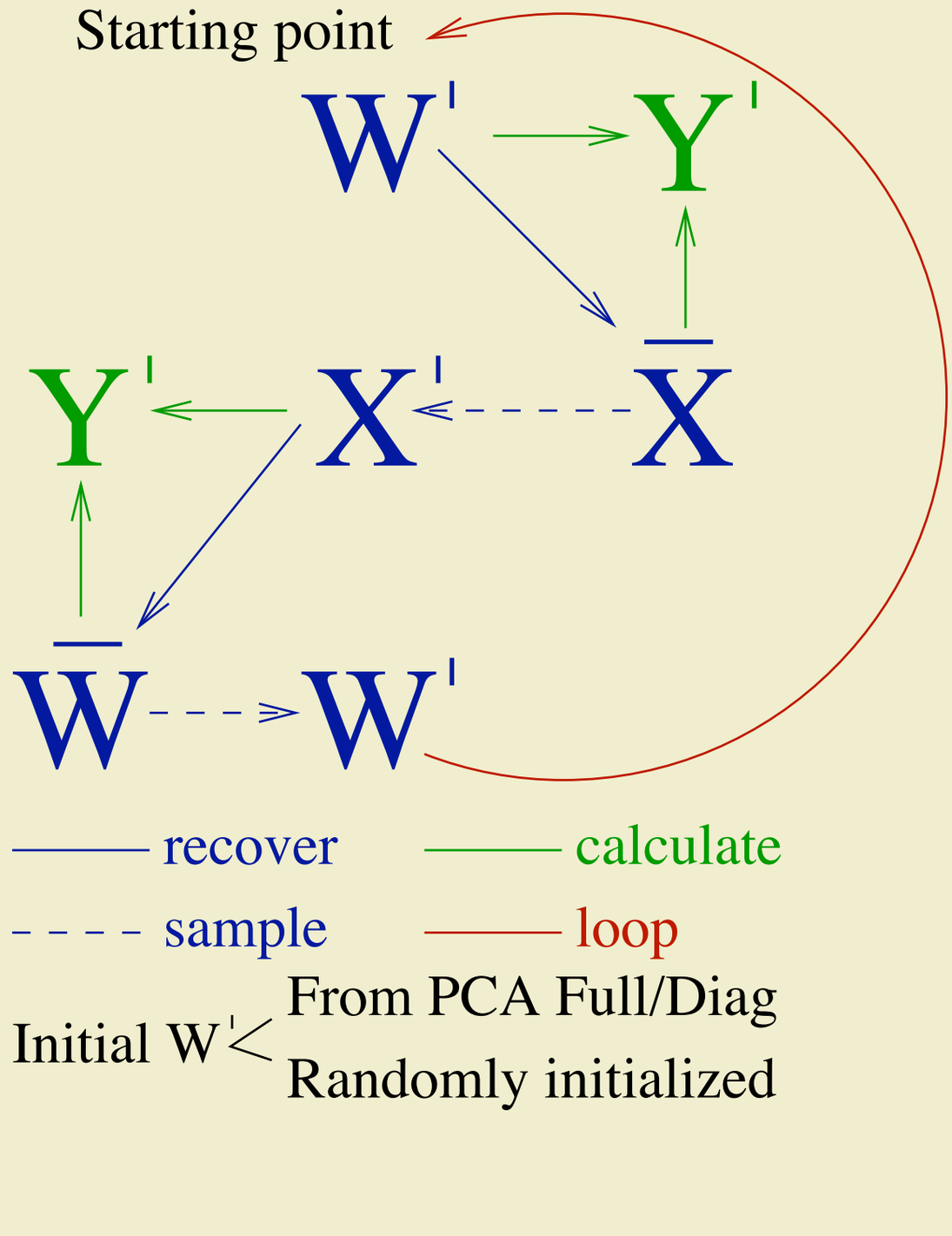
Variational Bayes



- Distribution with a fixed form is fitted to the **true distribution**
- The misfit between the two distributions is minimized iteratively

Gibbs sampling

- The distribution of W and m given Y and X has a simple form
- The distribution of X given Y, W and m has a simple form
- Draw random samples alternately from these

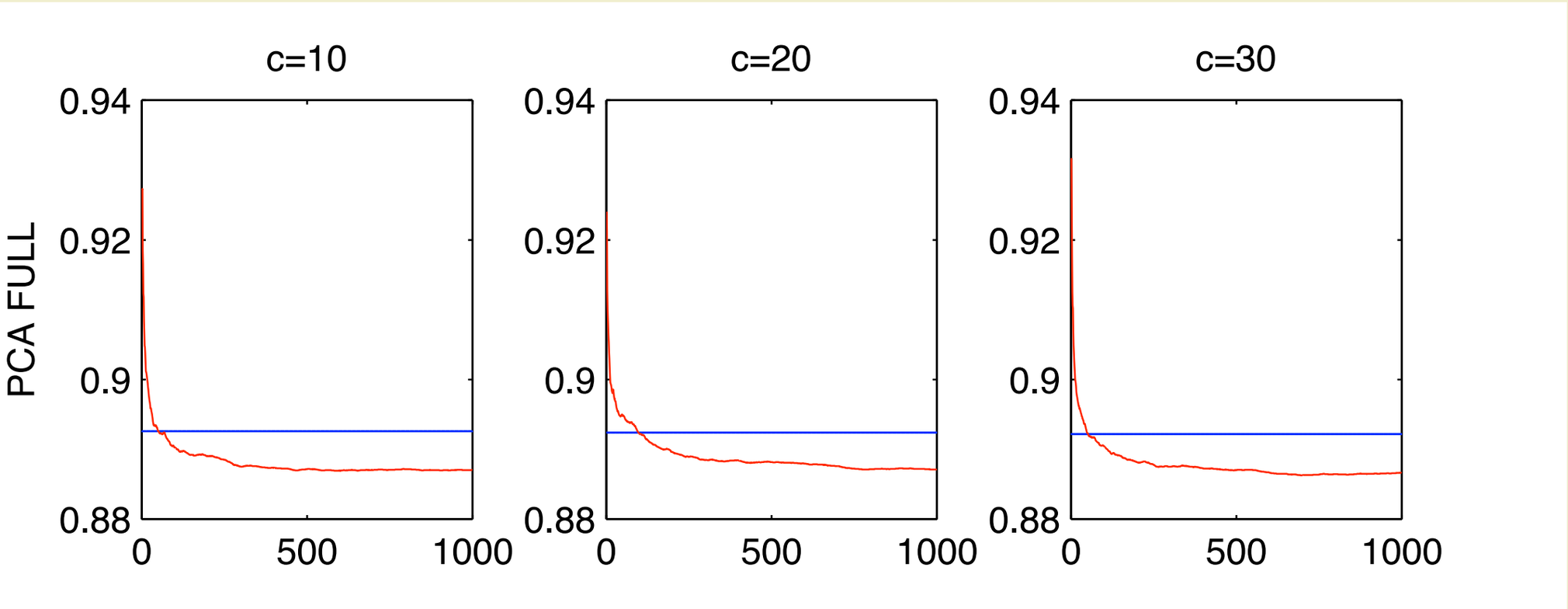


MovieLens data

- 100000 ratings from 1-5 given by 943 people to 1682 movies
- We used 95000 ratings for training and 5000 ratings for testing
- 94% missing values

Results with MovieLens data

- Sampling against variational Bayes with different number of components c=10,20,30
- RMS reconstruction error of test ratings



Conclusions

- When the data becomes very sparse, modelling uncertainty in PCA becomes important
- Sampling required about 100 samples to surpass variational Bayesian reconstructions