On Sparsity and Overcompleteness in Image Models

Richard Turner (turner@gatsby.ucl.ac.uk)

Pietro Berkes (berkes@gatsby.ucl.ac.uk)

Maneesh Sahani (maneesh@gatsby.ucl.ac.uk)

Gatsby Computational Neuroscience Unit, 12/06/2007

Motivation

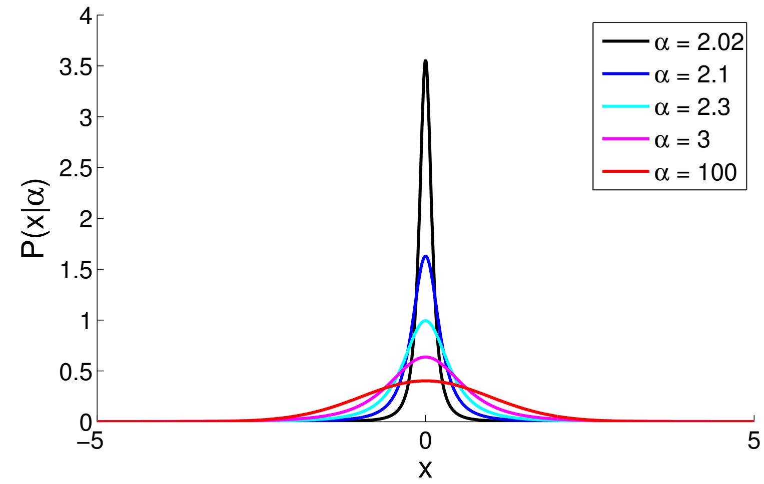
- **Sparse coding hypothesis**:
  - images are made from a **large library** of (**K**) structural-primitives
  - only a **small number** of primitives are active in a single image (sparseness α)
  - Early visual processing realises a representation of visual input in terms of these primitives

- Most work has **fixed** the library size (**K**) and sparseness and learned the structural primitives, validating results by dubious comparisons to biology

- GOAL: Discover what is the **optimal setting** of K and α from a **Bayesian perspective**

Sparse Coding Forward model

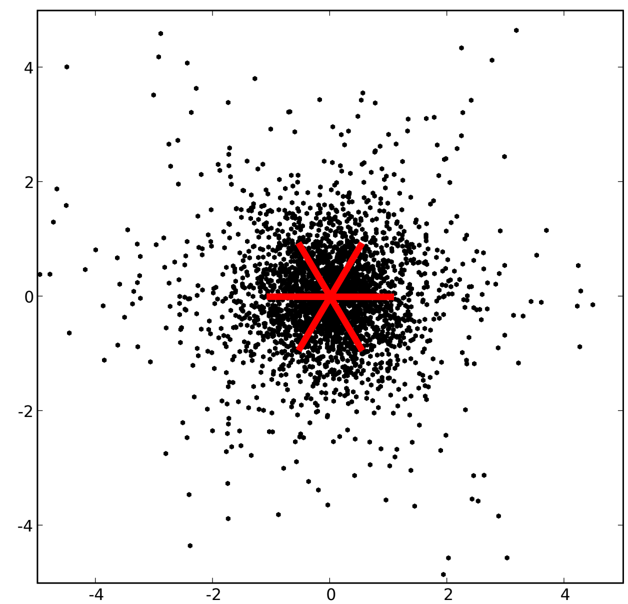
$$p(x_k|\alpha) \quad = \quad \frac{1}{Z}\left(\frac{1}{1+\frac{1}{\alpha}x_k^2}\right)^{-\frac{\alpha+1}{2}}$$



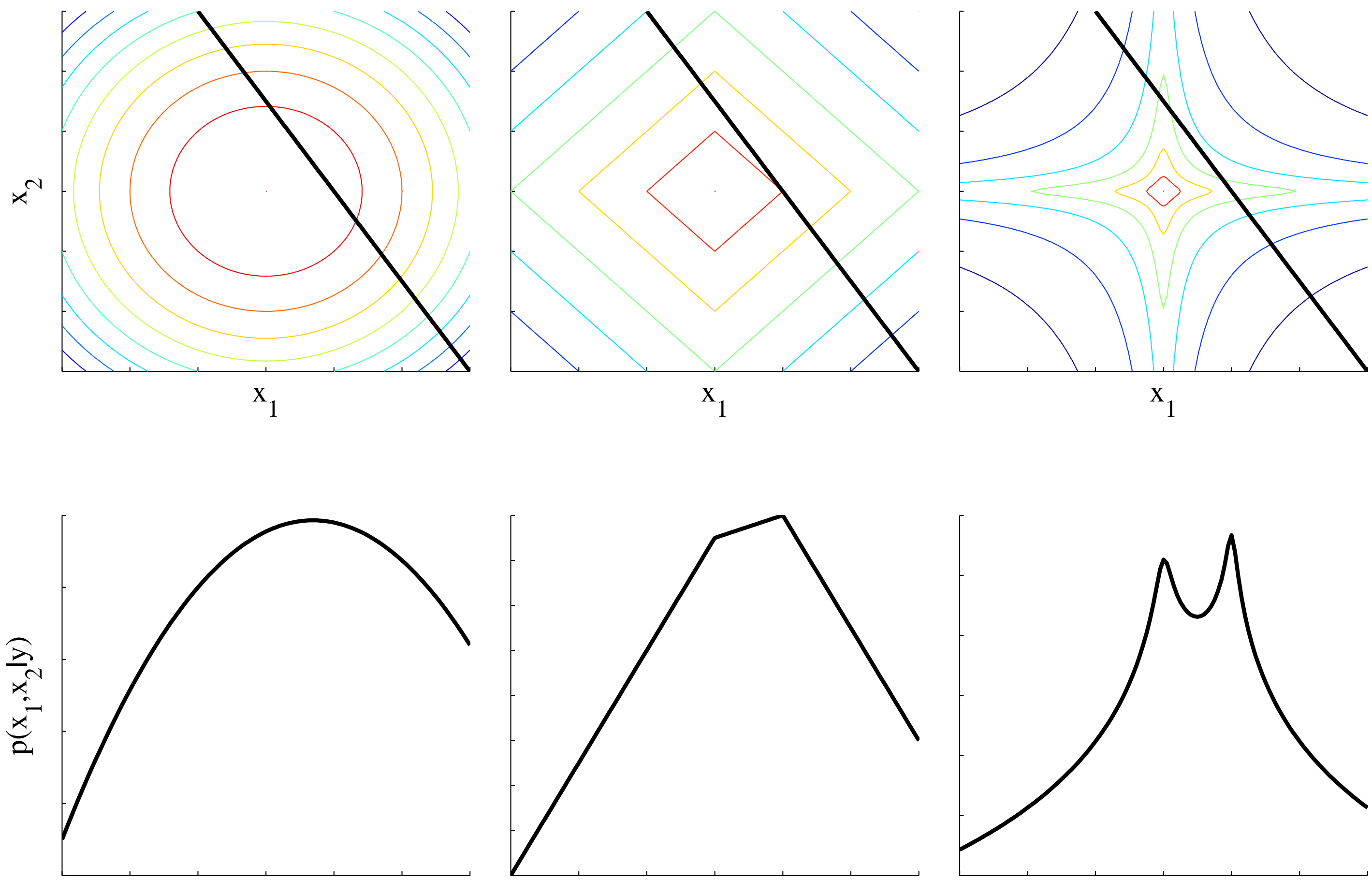
Sparse Coding Forward model

$$p(x_k|\alpha) = \frac{1}{Z}\left(\frac{1}{1+\frac{1}{\alpha}x_k^2}\right)^{-\frac{\alpha+1}{2}}$$

$$p(\mathbf{y}|\mathbf{x},\theta) = \mathrm{Norm}\left(\sum_k x_k \mathbf{w}_k, \sigma_y^2 I\right)$$



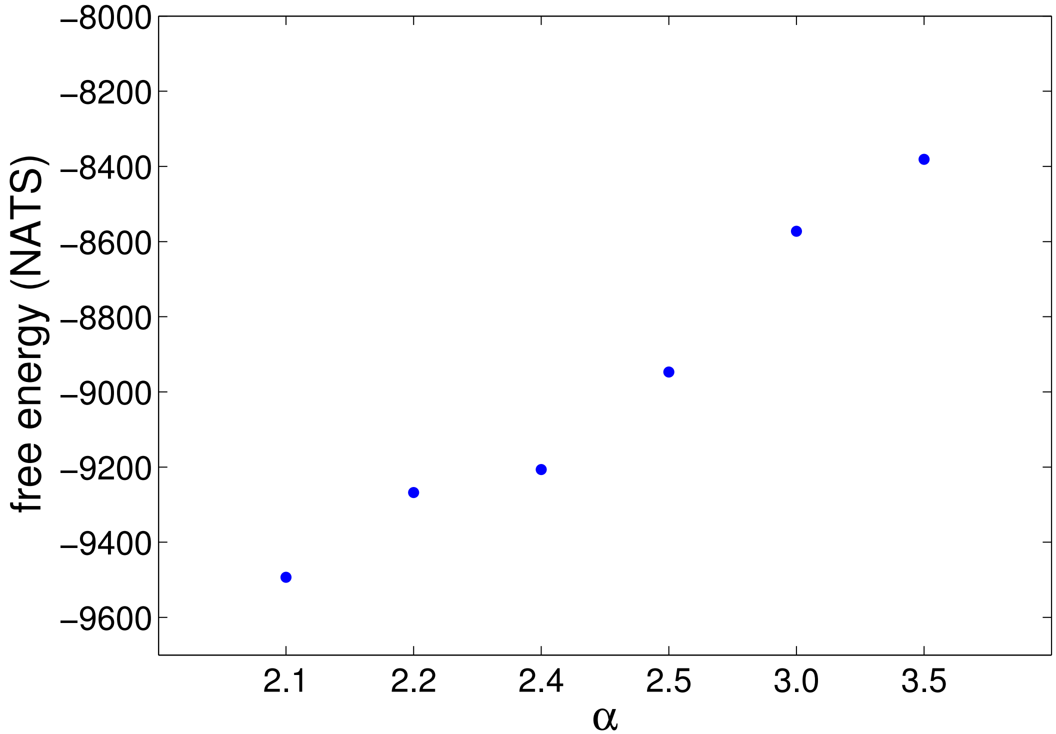
Why inference is hard: a toy example, D=1 K=2



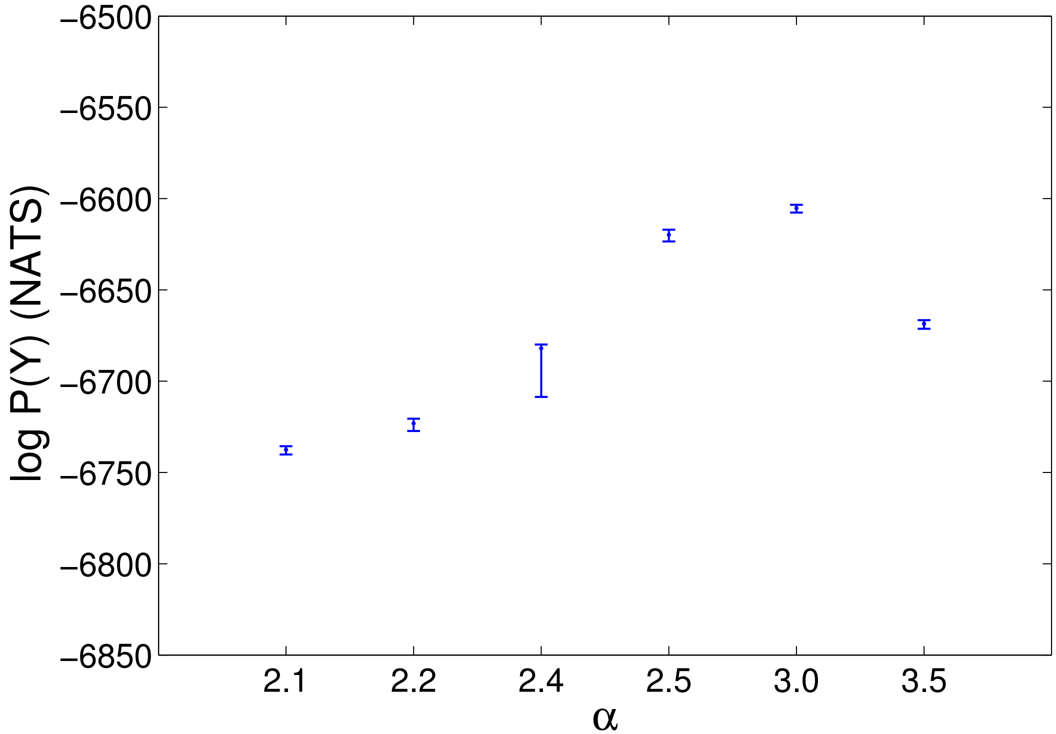
Methods

- **Ideally**: **Tile sparseness/overcompleteness**, learning models for each combination and compare the margingal likelihood.
  - But computationally **too slow** (a day per point)
- **Practically**: Learn the optimal over-completeness for a range of sparsenesses
  - **Automatic-relevance determination** (ARD) (turns a hard-model comparison problem into an easier parameter learning problem)
  - **Variational Bayesian** (VB) methods (v. fast)
- Evaluate the (approximate) marginal likelihood for each model
  - Cannot use VB as it is pathologically biased
  - Use a MCMC method called **Annealed Importance Sampling** (AIS)

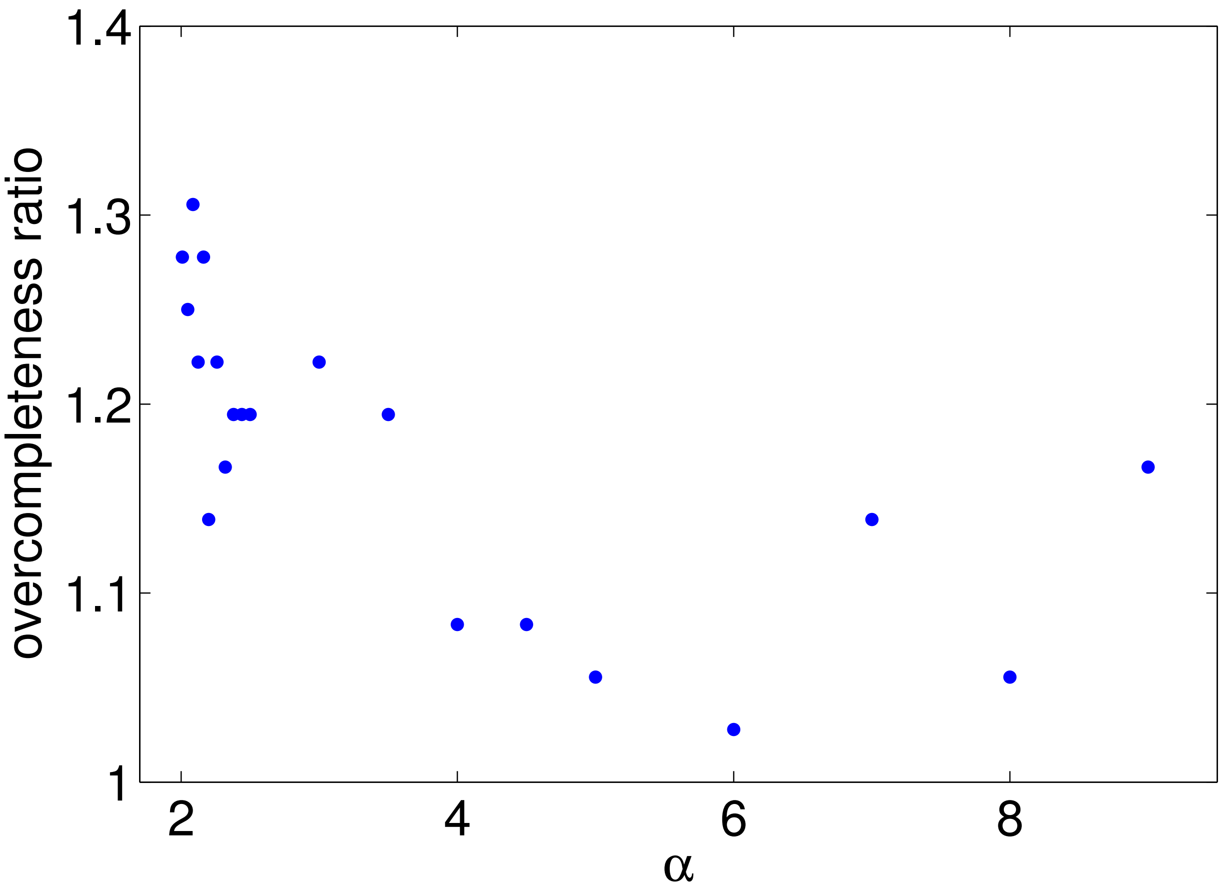
Results: VB in a toy model, α=2.5 D=2 K=3



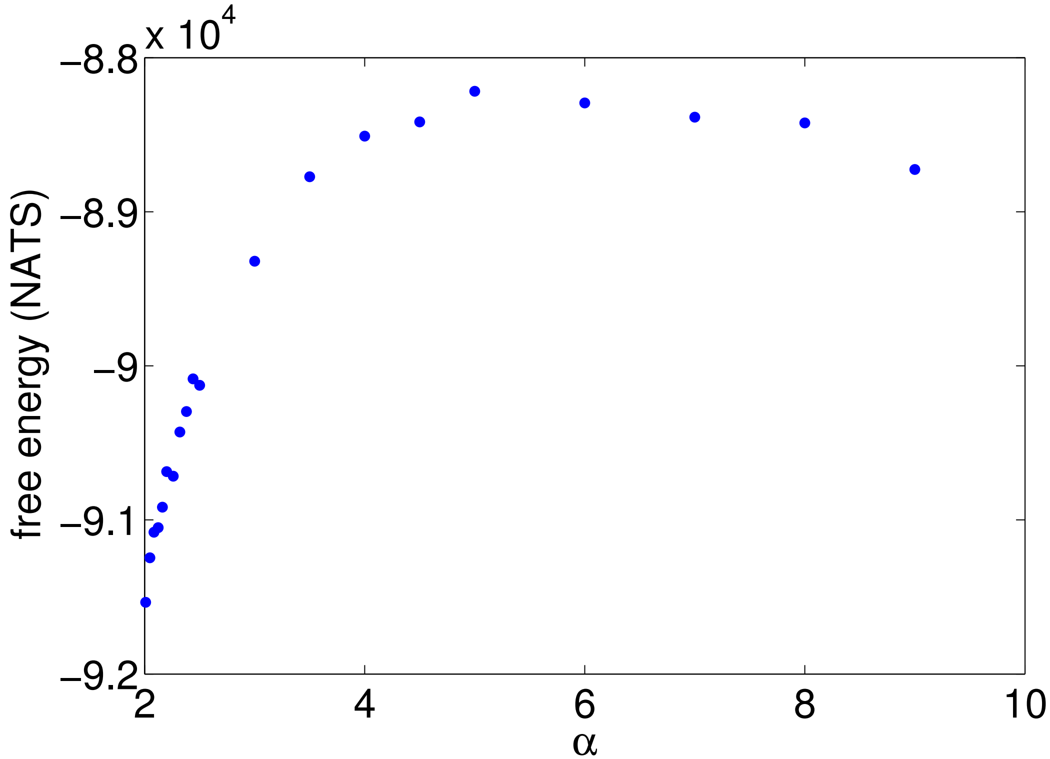
Results: AIS in a toy model, α=2.5 D=2 K=3



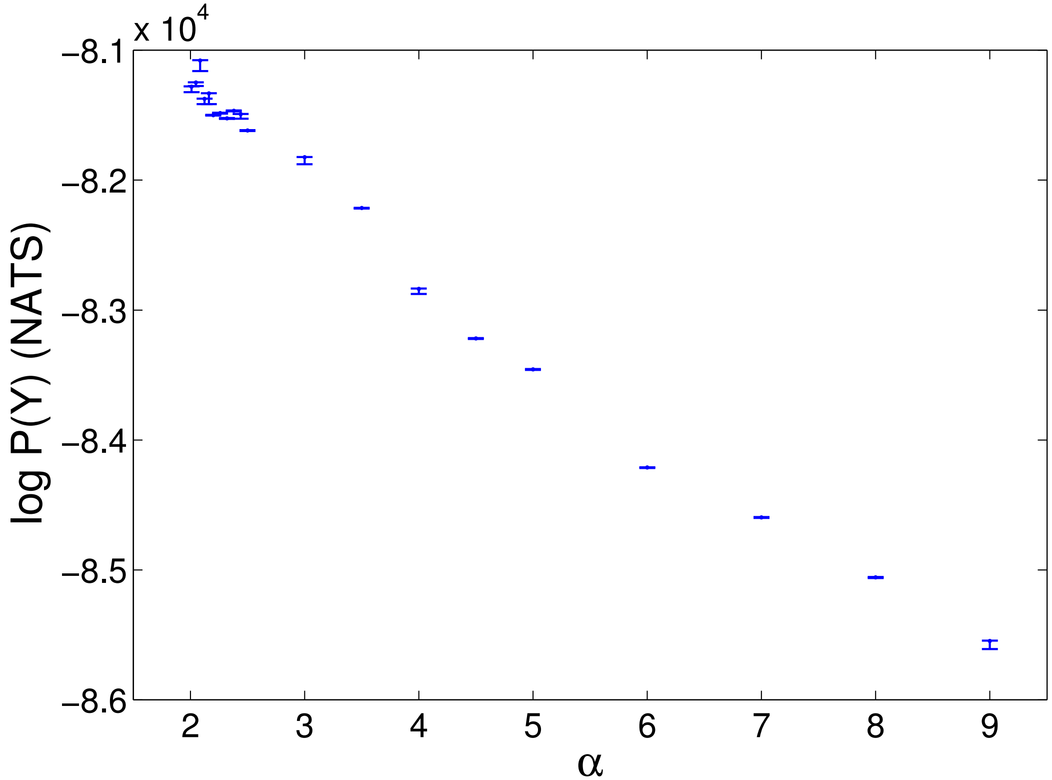
Results: VB for natural data, D=36 K=108



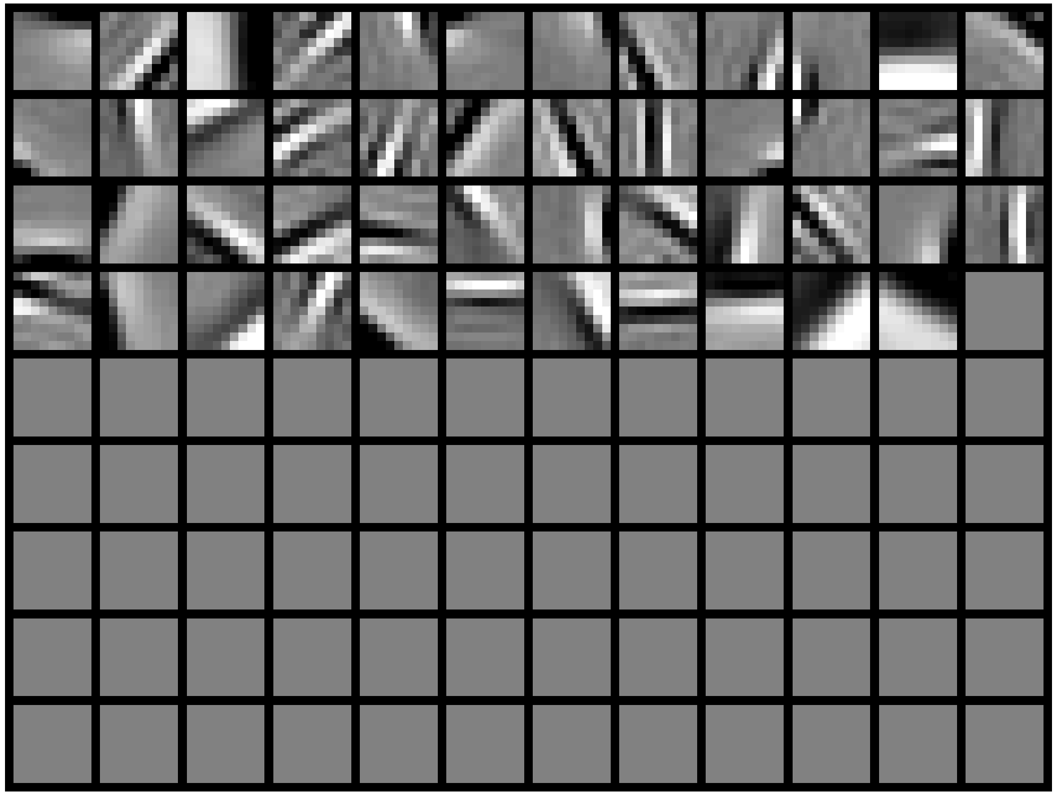
Results: VB for natural data, D=36 K=108



Results: AIS for natural data, D=36 K=108



Motivation



Conclusions

- The **optimal sparse coding model** is
  - indeed one that is **very sparse**
  - but only **modestly overcomplete**
- But, this is a statement **restricted to the Student-t prior** - should try other priors
- What effect do **non-linearities** have? (e.g. take a Gaussian Scale Mixture Model and check we can learn its dimensionality using the (wrong) linear model)
- How do the **structural primitives change** as we alter the sparsity and overcompleteness? Hypothesis: Auditory RFs will get more asymmetric.