Supplementary Material

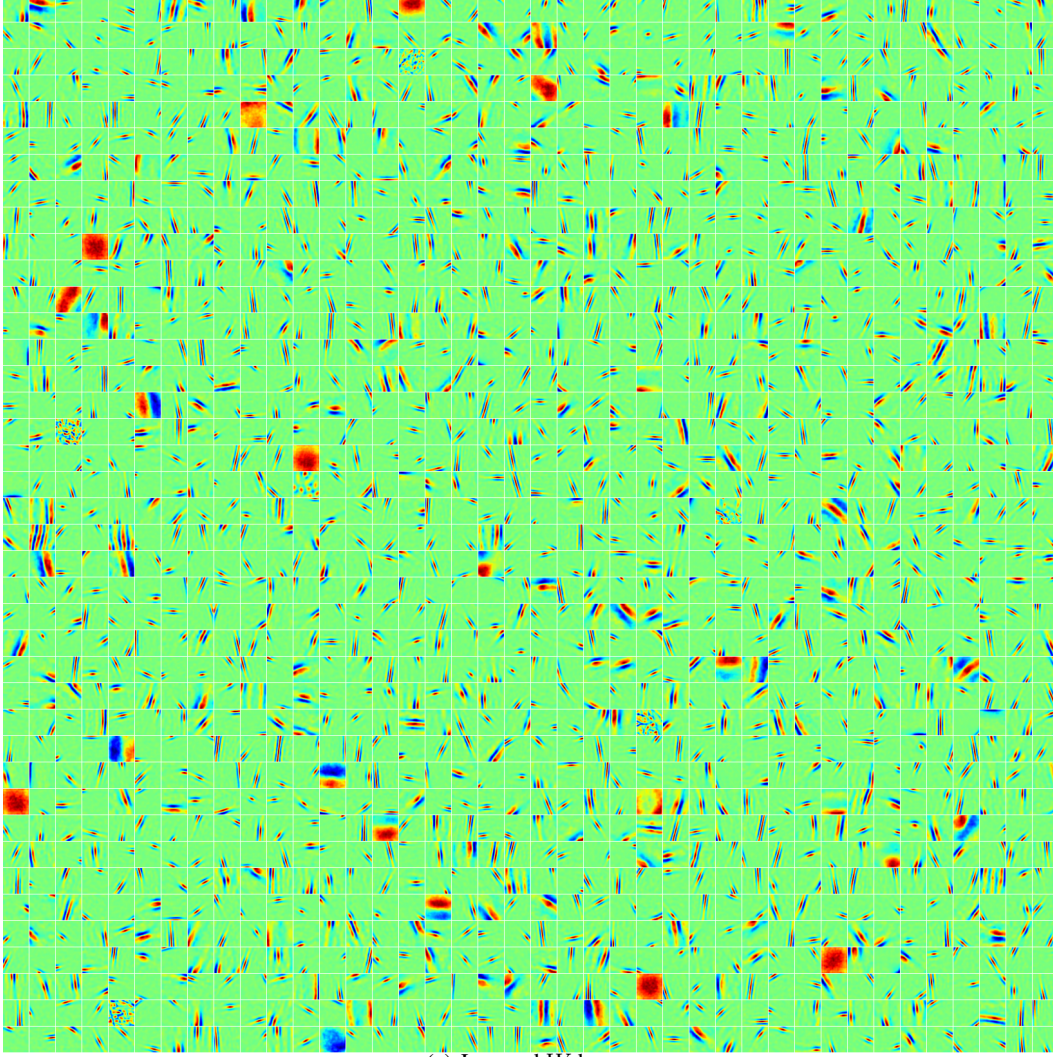
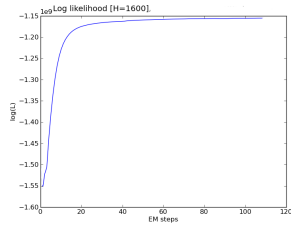
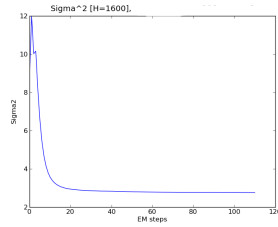
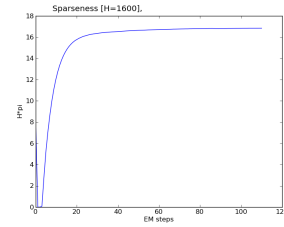
A BSC$^{\text{s+s}}$on 40×40 **Image patches**

(a) Learned W bases.

(b) Log-likelihood

(c) Learned σ^2.

(d) Learned $\pi H'$.

Figure 1: Results for BSC$^{\text{s+s}}$when running on $N = 500,000$ image patches of size $D = 40 \times 40 = 1600$. The number of hidden variables was set to $H = 1600$ and H' was set to 36. All other parameters were set as described in 4. (a) shows the learned basis functions $\vec{W}_h$, (b) shows the approximated log-likelihood throughout 100 EM steps and and corresponding learned data noise (c) and sparsity (d).

11

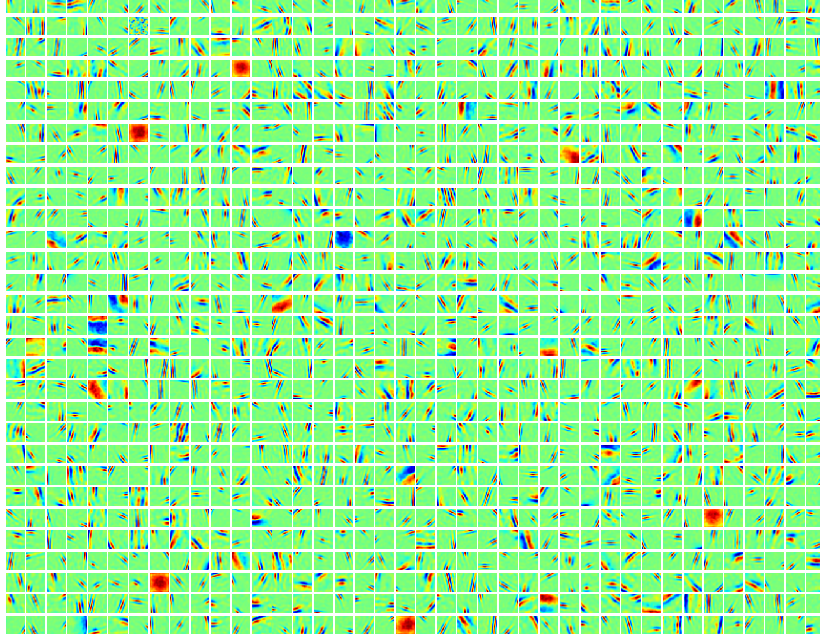
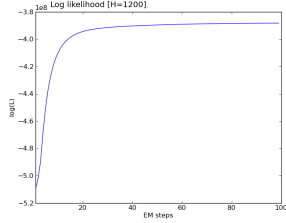
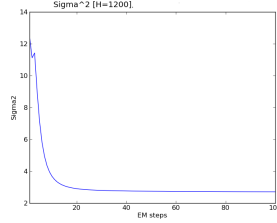
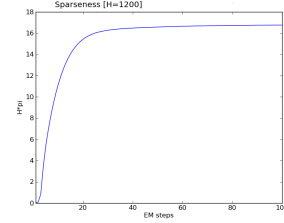
B BSC$^{\text{s+s}}$ on 32×32 Image patches

(a) Learned W bases.

(b) Log-likelihood

(c) Learned σ^2.

(d) Learned $\pi H'$.

Figure 2: Results for BSC$^{\text{s+s}}$ when running on $N = 200,000$ image patches of size $D = 32 \times 32 = 1024$. The number of hidden variables was set to $H = 1200$ and H' was set to 26. All other parameters were set as described in 4. (a) shows the learned basis functions $\vec{W}_h$, (b) shows the approximated log-likelihood throughout 100 EM steps and and corresponding learned data noise (c) and sparsity (d).

12