1 Gaussian Model

We consider a model in which $x \to y \to z$ forms a Markov chain, $p(x) \propto 1$ has a flat prior, and each conditional distribution is Gaussian. It follows that the posterior distribution has form

$$p(x, y \mid z) \triangleq \mathcal{N}(\mu^*, \Sigma), \tag{1}$$

where $\mu^* = (\mu_x^*, \mu_y^*)^\intercal$ and Σ has blocks Σ_{xx}, Σ_{xy}, and Σ_{yy}, and these means and covariances depend on the observation z. From Barber [eqn. 8.6.11] we know the conditionals have the form

$$p(x \mid z, y) = \mathcal{N}(x; \mu_x^* + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y^*), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}), \tag{2}$$
$$p(y \mid z, x) = \mathcal{N}(y; \mu_y^* + \Sigma_{yx}\Sigma_{xx}^{-1}(x - \mu_x^*), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}). \tag{3}$$

This gives the form of the CAVI variational factors

$$q^{(t+1)}(x) = \mathcal{N}(x; \mu_x^* + \Sigma_{xy}\Sigma_{yy}^{-1}(\widehat{\mu}_y^{(t)} - \mu_y^*), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}), \tag{4}$$
$$q^{(t+1)}(y) = \mathcal{N}(y; \mu_y^* + \Sigma_{yx}\Sigma_{xx}^{-1}(\widehat{\mu}_x^{(t)} - \mu_x^*), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}). \tag{5}$$

Where $\widehat{\mu}^{(t)}$ are the variational means at the t-th iteration. Hence

$$\widehat{\mu}_x^{(t+1)} - \mu_x^* = \Sigma_{xy}\Sigma_{yy}^{-1}(\widehat{\mu}_y^{(t)} - \mu_y^*) = \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}(\widehat{\mu}_x^{(t)} - \mu_x^*) \tag{6}$$
$$\widehat{\mu}_y^{(t+1)} - \mu_y^* = \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}(\widehat{\mu}_y^{(t)} - \mu_y^*) \tag{7}$$

Note that if λ is an operator of one of these matrices, i.e. $\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}v = \lambda v$, then

$$\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}(\Sigma_{xy}\Sigma_{yy}^{-1}v) = \lambda(\Sigma_{xy}\Sigma_{yy}^{-1}v),$$

and similarly in the other direction, so the spectra of these matrices coincide. Let $\gamma = \|\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\|_2 = \|\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\|_2$, so the rate of convergence of each variational mean is

$$\left\|\widehat{\mu}_x^{(t+1)} - \mu_x^*\right\|_2 \leq \gamma \left\|\widehat{\mu}_x^{(t)} - \mu_x^*\right\|_2 \tag{8}$$
$$\left\|\widehat{\mu}_y^{(t+1)} - \mu_y^*\right\|_2 \leq \gamma \left\|\widehat{\mu}_y^{(t)} - \mu_y^*\right\|_2 \tag{9}$$

Hence the rate of convergence for the whole algorithm is γ, since

$$\left\|(\widehat{\mu}_x^{(t+1)}, \widehat{\mu}_y^{(t+1)}) - (\mu_x^*, \mu_y^*)\right\|_2^2 = \left\|\widehat{\mu}_x^{(t+1)} - \mu_x^*\right\|_2^2 + \left\|\widehat{\mu}_y^{(t+1)} - \mu_y^*\right\|_2^2 \tag{10}$$
$$\leq \gamma^2 \left\|\widehat{\mu}_x^{(t)} - \mu_x^*\right\|_2^2 + \gamma^2 \left\|\widehat{\mu}_y^{(t)} - \mu_y^*\right\|_2^2 \tag{11}$$
$$= \gamma^2 \left\|(\widehat{\mu}_x^{(t)}, \widehat{\mu}_y^{(t)}) - (\mu_x^*, \mu_y^*)\right\|_2^2 \tag{12}$$

This rate γ matches the rate of convergence of the corresponding block Gibbs sampler [Sahu & Roberts, 1998, theorem 1].