

# Derivations for the CSM model

## 1 Definition of the model

A gestalt, a perceptual object, is characterised by a covariance component for the joint distribution of visual neural activity.

$$p(v \mid g) = \mathcal{N}(v; 0, C_v) \quad (1)$$

$$C_v = \sum_{k=1}^K g_k U_k^T U_k \quad (2)$$

where  $K$  is the fixed number of possible gestalts in the visual scene and  $g_k$  is the strength of the gestalt number  $k$ , coming from a  $K$ -dimensional Gamma prior distribution with shape and scale parameters  $\alpha_g$  and  $\theta_g$  controlling the sparsity of the prior.

$$p(g) = \text{Gam}(g; \alpha_g, \theta_g) \quad (3)$$

The global contrast of the image patch is encoded by a scalar variable  $z$ , also coming from a Gamma prior

$$p(z) = \text{Gam}(z; \alpha_z, \theta_z) \quad (4)$$

The pixel intensities are generated from the neural activity through a set of linear projective field models, possibly Gabor filters,  $A$ , scaled by the contrast and adding some independent observational noise.

$$p(x \mid v, z) = \mathcal{N}(x; zAv, \sigma_x I) \quad (5)$$

## 2 Likelihoods

### 2.1 Likelihood of $g$ and $z$

by intuition:

$$p(x \mid z, g) = \mathcal{N}(x; 0, \sigma_x I + z^2 A \left( \sum_{k=1}^K g_k U_k^T U_k \right) A^T) \quad (6)$$

by algebraic derivation (see Seq. 4.5):

$$p(x | z, g) = \int_{-\infty}^{\infty} p(x, v | z, g) dv = \frac{1}{z^{D_v} \sqrt{\det(A^T A)}} \mathcal{N}\left(\frac{1}{z} A^+ x; 0, \frac{\sigma_x}{z^2} (A^T A)^{-1} + \sum_{k=1}^K g_k U_k^T U_k\right) \quad (7)$$

$$f(x, z) \equiv \frac{1}{z} A^+ x, \quad h(z) \equiv \frac{1}{z^{D_v}} \quad (8)$$

$$C(z, g) \equiv \frac{\sigma_x}{z^2} (A^T A)^{-1} + \sum_{k=1}^K g_k U_k^T U_k \quad (9)$$

$$p(x | z, g) = \frac{1}{\sqrt{\det(A^T A)}} h(z) \mathcal{N}(f(x, z); 0, C(z, g)) \quad (10)$$

## 2.2 Log-likelihood of the parameters

All parameters of the model consist of

$$\zeta = \{\sigma_x, A, U_{1..K}, \alpha_g, \theta_g, \alpha_z, \theta_z\} \quad (11)$$

for  $n$  exchangeable observations of  $x$ , the likelihood looks like this

$$p(X | \zeta) = \prod_{n=1}^N p(x_n | \zeta) = \prod_{n=1}^N \iint_{-\infty}^{\infty} p(x_n | z, g) p(g) p(z) dg dz \quad (12)$$

approximation of the integrals by samples from the priors  $p(g)$  and  $p(z)$

$$p(X | \zeta) \approx \prod_{n=1}^N \frac{1}{L} \sum_{l=1}^L p(x_n | z^l, g^l) \quad (13)$$

using Eq. 10

$$p(X | \zeta) \approx \left( L \sqrt{\det(A^T A)} \right)^{-N} \prod_{n=1}^N \sum_{l=1}^L h(z^l) \mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l)) \quad (14)$$

$$\log p(X | \zeta) \approx -N(\log L + \frac{1}{2} \log(\det(A^T A))) + \sum_{n=1}^N \log \left[ \sum_{l=1}^L h(z^l) \mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l)) \right] \quad (15)$$

$$h^l \equiv h(z^l), \quad f_n^l \equiv f(x_n, z^l), \quad C^l \equiv C(z^l, g^l) \quad (16)$$

$$\mathcal{L}_n^l \equiv h^l \mathcal{N}(f_n^l; 0, C^l), \quad \mathcal{L}_n \equiv \sum_{l=1}^L \mathcal{L}_n^l \quad (17)$$

$$\log p(X | \zeta) \approx -N(\log L + \frac{1}{2} \log(\det(A^T A))) + \sum_{n=1}^N \log \mathcal{L}_n \quad (18)$$

An equivalent way to write this based on Eq. 6 is the following

$$p(X | \zeta) \approx L^{-N} \prod_{n=1}^N \sum_{l=1}^L \mathcal{N}(x_n; 0, \sigma_x I + z^{l2} A C_v^l A^T) \quad (19)$$

$$C_x^l \equiv \sigma_x I + z^{l2} A C_v^l A^T \quad (20)$$

$$\log p(X | \zeta) \approx -N \log L + \sum_{n=1}^N \log \left[ \sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l) \right] \quad (21)$$

### 2.2.1 Derivative w.r.t. $U_{1...K}$

Using Eq. 18

$$\begin{aligned} \frac{\partial \log p(X | \zeta)}{\partial [U_k]_{i,j}} &\approx \sum_{n=1}^N \frac{\partial \log \mathcal{L}_n}{\partial [U_k]_{i,j}} = \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \frac{\partial \mathcal{L}_n}{\partial [U_k]_{i,j}} = \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L \frac{\partial \mathcal{L}_n^l}{\partial [U_k]_{i,j}} = \\ &= \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L \text{Tr} \left[ \frac{\partial \mathcal{L}_n^l}{\partial C^l} \frac{\partial C^l}{\partial [U_k]_{i,j}} \right] \end{aligned} \quad (22)$$

the derivatives in this formula are the following

$$\begin{aligned} \frac{\partial \mathcal{L}_n^l}{\partial C^l} &= h^l \frac{\partial}{\partial C^l} \mathcal{N}(f_n^l; 0, C^l) = h^l \mathcal{N}(f_n^l; 0, C^l) \frac{\partial}{\partial C^l} \log \mathcal{N}(f_n^l; 0, C^l) = \\ &= -\frac{h^l}{2} \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \end{aligned} \quad (23)$$

$$\frac{\partial C(z, g)}{\partial [U_k]_{i,j}} = g_k \frac{\partial (U_k^T U_k)}{\partial [U_k]_{i,j}} = g_k (U_k^T J^{ij} + J^{ji} U_k) \equiv g_k \hat{U}_k^{ij} \quad (24)$$

substituting back to the derivative

$$\begin{aligned} \frac{\partial \log p(X | \zeta)}{\partial [U_k]_{i,j}} &\approx \\ &\approx -\frac{1}{2} \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) \text{Tr} \left[ [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \hat{U}_k^{ij} \right] = \\ &= -\frac{1}{2} \text{Tr} \left[ \sum_{n=1}^N \left( \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \right) \hat{U}_k^{ij} \right] \end{aligned} \quad (25)$$

The regularities of the  $\hat{U}_k$  matrices allow us to replace the trace with a much more efficient computation:

$$M_k = -\frac{1}{2} \sum_{n=1}^N \left( \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \right) \quad (26)$$

$$\frac{\partial \log p(X | \zeta)}{\partial [U_k]_{i,j}} \approx \text{Tr} [M_k \hat{U}_k^{ij}] = \sum_{a=1}^{Dv} [M_k]_{j,a} [U_k]_{i,a} + [M_k]_{a,j} [U_k]_{i,a} \quad (27)$$

also, by substituting back Eq. 24 we get

$$\text{Tr} [M_k \hat{U}_k^{ij}] = \text{Tr} [M_k U_k^T J^{ij}] + \text{Tr} [M_k J^{ji} U_k] \quad (28)$$

that is by equivalences of single-entry matrices and noting that  $M_k$  are symmetric

$$\text{Tr} [M_k \hat{U}_k^{ij}] = [U_k M_k^T]_{i,j} + [U_k M_k]_{i,j} = 2 [U_k M_k]_{i,j} \quad (29)$$

yielding

$$\frac{\partial \log p(X | \zeta)}{\partial U_k} \approx -U_k \sum_{n=1}^N \left( \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \right) \quad (30)$$

### 2.2.2 Derivative w.r.t. $\sigma_x$

Using Eq. 21

$$\begin{aligned} \frac{\partial \log p(X | \zeta)}{\partial \sigma_x} &\approx \sum_{n=1}^N \frac{\partial}{\partial \sigma_x} \log \left[ \sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l) \right] = \\ &= \sum_{n=1}^N \frac{1}{\sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l)} \sum_{l=1}^L \frac{\partial}{\partial \sigma_x} \mathcal{N}(x_n; 0, C_x^l) = \\ &= \sum_{n=1}^N \frac{1}{\sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l)} \sum_{l=1}^L \text{Tr} \left[ \frac{\partial}{\partial C_x^l} \mathcal{N}(x_n; 0, C_x^l) \frac{\partial C_x^l}{\partial \sigma_x} \right] \end{aligned} \quad (31)$$

the derivatives in this formula are the following

$$\begin{aligned} \frac{\partial}{\partial C_x^l} \mathcal{N}(x_n; 0, C_x^l) &= \mathcal{N}(x_n; 0, C_x^l) \frac{\partial}{\partial C^l} \log \mathcal{N}(x_n; 0, C_x^l) = \\ &= -\frac{1}{2} \mathcal{N}(x_n; 0, C_x^l) [(C_x^l)^{-1} - (C_x^l)^{-1} x_n x_n^T (C_x^l)^{-1}] \end{aligned} \quad (32)$$

$$\frac{\partial C_x^l}{\partial \sigma_x} = I \quad (33)$$

substituting back to the derivative

$$\begin{aligned} & \frac{\partial \log p(X | \zeta)}{\partial \sigma_x} \approx \\ & \approx -\frac{1}{2} \sum_{n=1}^N \frac{1}{\sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l)} \sum_{l=1}^L \mathcal{N}(x_n; 0, C_x^l) \text{Tr} [(C_x^l)^{-1} - (C_x^l)^{-1} x_n x_n^T (C_x^l)^{-1}] \end{aligned} \quad (34)$$

### 2.3 Complete-data log-likelihood

$$p(V, G, Z, X | \zeta) = \prod_{n=1}^N p(x_n | v_n, z_n) p(v_n | g_n) p(g_n) p(z_n) \quad (35)$$

the logarithm of this will be

$$\begin{aligned} \log p(V, G, Z, X | \zeta) &= \\ &= \sum_{n=1}^N [\log p(g_n) + \log p(z_n) + \log p(x_n | v_n) + \log p(v_n | g_n)] = \\ &= N(\log p(g) + \log p(z)) + \sum_{n=1}^N \log p(x_n | v_n) + \log p(v_n | g_n) = \\ &= c + \sum_{n=1}^N \log p(v_n | g_n) \end{aligned} \quad (36)$$

where  $c$  is constant with respect to the parameters  $U_{1...K}$ .

#### 2.3.1 Expectation w.r.t. the posterior

$$\mathcal{L} = \iiint_{-\infty}^{\infty} p(V, G, Z | X) \log p(V, G, Z, X | \zeta) dV dG dZ. \quad (37)$$

We can approximate this integral by averaging over  $L$  samples from the full posterior, separately for each observation  $x_n$ , as defined in Seq. 3.1. As we will seek the values of the Cholesky components  $U_{1...K}$  that maximise this integral, we can discard each term not depending on these parameters, only leaving the term of the form  $p(v | g)$ . This way we arrive to the following expression

$$\begin{aligned} \mathcal{L} &\sim \sum_{n=1}^N \frac{1}{L} \sum_{l=1}^L -\frac{1}{2} \left[ \log \left( \det \left( C_v^{(l,n)} \right) \right) + v^{(l,n)T} (C_v^{(l,n)})^{-1} v^{(l,n)} \right] = \\ &= -\frac{1}{2L} \sum_{m=1}^{NL} \left[ \log \left( \det \left( C_v^m \right) \right) + v^{mT} (C_v^m)^{-1} v^m \right] \end{aligned} \quad (38)$$

noting that the double summation over  $L$  samples over all  $N$  observations always happens on the same terms, so we can substitute it with a single sum that iterates over the full sample set.

### 2.3.2 Derivative w.r.t. $U_{1...K}$

Using the derivative of  $C_v$  with respect to  $U_k^{ij}$  as defined in Eq. 24, by the chain rule, the derivative of  $\mathcal{L}$  according to an element of  $U_k$  looks like this

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} &= -\frac{1}{2L} \sum_{m=1}^{NL} \text{Tr} \left[ \frac{\partial \mathcal{L}^m}{\partial C_v^m} \frac{\partial C_v^m}{\partial [U_k]_{i,j}} \right] = \\ &- \frac{1}{2L} \sum_{m=1}^{LN} \text{Tr} \left[ \left[ (C_v^m)^{-1} - (C_v^m)^{-1} v^m v^{mT} (C_v^m)^{-1} \right] g_k^m \hat{U}_k^{ij} \right] = \\ &- \frac{1}{2L} \text{Tr} \left[ \sum_{m=1}^{LN} g_k^m \left[ (C_v^m)^{-1} - (C_v^m)^{-1} (v^m v^{mT}) (C_v^m)^{-1} \right] \hat{U}_k^{ij} \right] \end{aligned} \quad (39)$$

The regularities of the  $\hat{U}_k$  matrices allow us to replace the trace with a much more efficient computation:

$$M_k = -\frac{1}{2L} \left[ \sum_{m=1}^{LN} g_k^m \left[ (C_v^m)^{-1} - (C_v^m)^{-1} (v^m v^{mT}) (C_v^m)^{-1} \right] \right] \quad (40)$$

$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \sum_{a=1}^{Dv} [M_K]_{j,a} [U_k]_{i,a} + [M_k]_{a,j} [U_k]_{i,a} \quad (41)$$

## 3 Posteriors

### 3.1 Full posterior

$$p(v, g, z | x) = p(x | v, g, z) p(v | g) p(g) p(z) \frac{1}{p(x)} \sim p(x | v, z) p(v | g) p(g) p(z) \quad (42)$$

so the log-posterior will be the following, up to an additive constant, using Gamma priors over  $g$  and  $z$  defined by shape and scale parameters:

$$\begin{aligned} \log p(v, g, z | x) &\sim \log p(x | v, z) + \log p(v | g) + \log p(g) + \log p(z) = \\ &= \log \mathcal{N}(x; zAv, \sigma_x I) + \log \mathcal{N}(v; 0, C_v) + \log \text{Gam}(g; \alpha_g, \theta_g) + \log \text{Gam}(z; \alpha_z, \theta_z) \end{aligned} \quad (43)$$

using the logarithms of the used pdfs from Sec. 4.2 and discarding all terms not dependent on any of the three variables we get

$$\begin{aligned}
\log p(v, g, z | x) &\sim -\frac{1}{2\sigma_x} (x - zAv)^T (x - zAv) - \\
&\quad -\frac{1}{2} [\log(\det(C_v)) + v^T C_v^{-1} v] + \\
&\quad + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z}
\end{aligned} \tag{44}$$

rearranging the first quadratic term according to Eq. 89 and discarding the term not dependent on  $v$  yields

$$\begin{aligned}
\log p(v, g, z | x) &\sim -\frac{z}{2\sigma_x} (zv^T A^T Av - 2x^T Av) - \\
&\quad -\frac{1}{2} [\log(\det(C_v)) + v^T C_v^{-1} v] + \\
&\quad + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + \\
&\quad + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z}
\end{aligned} \tag{45}$$

### 3.1.1 Derivative w.r.t. $v$

$$\log p(v, g, z | x) \sim -\frac{1}{2} \left[ \frac{z^2}{\sigma_x} v^T A^T Av - \frac{2z}{\sigma_x} x^T Av + v^T C_v^{-1} v \right] + f_1(g, z) \tag{46}$$

lumping the two quadratic forms together

$$\log p(v, g, z | x) \sim \frac{z}{\sigma_x} x^T Av - \frac{1}{2} v^T \left[ \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right] v + f_1(g, z) \tag{47}$$

Taking the derivative using Eq. 76 and 77 we get

$$\frac{\partial}{\partial v} \log p(v, g, z | x) = \frac{z}{\sigma_x} A^T x - \left[ \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right] v \tag{48}$$

### 3.1.2 Derivative w.r.t. $g$

$$\begin{aligned}
\log p(v, g, z | x) &\sim -\frac{1}{2} [\log \det(C_v) + v^T C_v^{-1} v] + \\
&\quad + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + f_2(v, z)
\end{aligned} \tag{49}$$

Taking the derivative w.r.t. a single  $g_i$  using Eq. 80 we get

$$\begin{aligned} \frac{\partial}{\partial g_i} \log p(v, g, z | x) = & -\frac{1}{2} \text{Tr} \left[ \frac{\partial}{\partial C_v} [\log \det(C_v) + v^T C_v^{-1} v] \frac{\partial C_v}{\partial g_i} \right] + \\ & + \frac{\partial}{\partial g_i} \left[ (\alpha_g - 1) \log(g_i) - \frac{g_i}{\theta_g} \right] \end{aligned} \quad (50)$$

using Eq. 79, 78 and 81 we arrive to

$$\frac{\partial}{\partial g_i} \log p(v, g, z | x) = -\frac{1}{2} \text{Tr} \left[ [C_v^{-1} - C_v^{-1} v v^T C_v^{-1}] C_i \right] + \frac{\alpha_g - 1}{g_i} - \frac{1}{\theta_g} \quad (51)$$

### 3.1.3 Derivative w.r.t. $z$

$$\log p(v, g, z | x) \sim -\frac{z}{2\sigma_x} (z v^T A^T A v - 2x^T A v) + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} + f_3(g, v) \quad (52)$$

$$\frac{\partial}{\partial z} \log p(v, g, z | x) = \frac{1}{\sigma_x} [x^T A v - z v^T A^T A v] + \frac{\alpha_z - 1}{z} - \frac{1}{\theta_z} \quad (53)$$

## 3.2 Conditional posteriors

### 3.2.1 Conditional posterior of $v$

$$p(v | x, g, z) = \frac{p(x | v, z, g) p(v | z, g)}{p(x | z, g)} = \frac{\mathcal{N}(x; z A v, \sigma_x I) \mathcal{N}(v; 0, C_v)}{\int_{-\infty}^{\infty} \mathcal{N}(x; z A v, \sigma_x I) \mathcal{N}(v; 0, C_v) dv} \quad (54)$$

the product of two Gaussians in the numerator of Eq. 54 can also be written as a Gaussian over  $v$  as in Seq. 4.4:

$$\mathcal{N}(x; z A v, \sigma_x I) \mathcal{N}(v; 0, C_v) = c \mathcal{N}(v; \mu_{post}, C_{post}) \quad (55)$$

The denominator of Eq. 54 is the integral of this formula, which evaluates to  $c$ , as the Gaussian integrates to one. This cancels the constant in the numerator, making the conditional posterior equal to the combined Gaussian over  $v$ , which, after expanding  $\mu_{post}$  and  $C_{post}$ , is

$$p(v | x, g, z) = \mathcal{N} \left( v; \frac{z}{\sigma_x} \left( \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right)^{-1} A^T x, \left( \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right)^{-1} \right) \quad (56)$$

### 3.2.2 Conditional posterior of $g$

$$p(g | X, V, z) = \frac{p(X | g, V, z) p(g | V, z)}{p(X | V, z)} = \frac{p(V | g) p(g)}{p(V)} \quad (57)$$

taking the logarithm and discarding constant terms



$$\log p(g \mid X, V) \sim -\frac{1}{2} [\log(\det(C_v)) + v^T C_v^{-1} v] + \log p(g) \quad (58)$$

The unnormalised conditionals of single elements of  $g$ , assuming an independent prior look as follows

$$\begin{aligned} \log p(g_j \mid g_{\neg j}, X, V) &= \frac{p(V \mid g_j, g_{\neg j}, X) p(g_j \mid g_{\neg j}, X)}{p(V \mid g_{\neg j}, X)} = \\ &= \frac{p(V \mid g) p(g_j)}{p(V \mid g_{\neg j})} \sim p(V \mid g) p(g_j) \end{aligned} \quad (59)$$

### 3.2.3 Conditional posterior of $z$

$$p(z \mid X, V, g) = \frac{p(X \mid g, z, V) p(z \mid V, g)}{p(X \mid V, g)} \sim p(X \mid z, V) p(z) \quad (60)$$

the log-posterior being

$$\log p(z \mid X, V) \sim -\frac{1}{2} \left[ D_x \log(\sigma_x) + \frac{1}{\sigma_x} (x - zAv)^T (x - zAv) \right] + \log p(z) \quad (61)$$

## 3.3 Marginal posteriors

### 3.3.1 Marginal posterior of $g$ and $z$

$$p(g, z \mid x) \sim p(x \mid g, z) p(g) p(z) \quad (62)$$

from Eq. 6

$$\begin{aligned} \log p(g, z \mid x) &\sim -\frac{1}{2} [\log \det(C_x) + x^T C_x^{-1} x] + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} + \\ &\quad + (\alpha_g - 1) \sum_{k=1}^K \log(g_k) - \frac{1}{\theta_g} \sum_{k=1}^K g_k \end{aligned} \quad (63)$$

$$C_x = \sigma_x I + z^2 A \left( \sum_{k=1}^K g_k U_k^T U_k \right) A^T \quad (64)$$

the derivative w.r.t  $g_k$

$$\begin{aligned} \frac{\partial}{\partial g_k} \log p(g, z \mid x) &= \text{Tr} \left[ \frac{\partial}{\partial C_x} \log p(g, z \mid x) \frac{\partial C_x}{\partial g_k} \right] + \frac{\alpha_g - 1}{g_k} - \frac{1}{\theta_g} = \\ &= -\frac{z^2}{2} \text{Tr} \left[ [C_x^{-1} - C_x^{-1} x x^T C_x^{-1}] A U_k^T U_k A^T \right] + \frac{\alpha_g - 1}{g_k} - \frac{1}{\theta_g} \end{aligned} \quad (65)$$

the derivative w.r.t  $z$

$$\begin{aligned} \frac{\partial}{\partial z} \log p(g, z | x) &= \text{Tr} \left[ \frac{\partial}{\partial C_x} \log p(g, z | x) \frac{\partial C_x}{\partial z} \right] + \frac{\alpha_z - 1}{z} - \frac{1}{\theta_z} = \\ &= -z \text{Tr} \left[ [C_x^{-1} - C_x^{-1} x x^T C_x^{-1}] A C_v A^T \right] + \frac{\alpha_z - 1}{z} - \frac{1}{\theta_z} \end{aligned} \quad (66)$$

### 3.3.2 Marginal posterior of $g$

A maximum a posterior estimate of  $g$  can be given as follows

$$g_{MAP} = \arg \max_g p(g | x) = \arg \max_g \frac{p(x | g)p(g)}{p(x)} = \arg \max_g p(x | g)p(g) \quad (67)$$

$$p(x | g) = \int_{-\infty}^{\infty} p(x | z, g)p(z)dz \approx \frac{1}{L} \sum_{l=1}^L p(x | g, z^l) \quad (68)$$

### 3.3.3 Marginal posterior of $v$

$$p(v | x) = \iint_{-\infty}^{\infty} p(v | x, g, z)p(g, z | x)dgdz \quad (69)$$

$$p(v | x) \approx \frac{1}{L} \sum_{l=1}^L p(v | x, g^l, z^l), \quad g^l, z^l \sim p(g, z | x) \quad (70)$$

where  $p(v | x, g, z)$  is given by Eq. 56, so we approximate the marginal posterior with a finite mixture of Gaussians, for which the covariance is given in the following form

$$C_{v|x} \approx \frac{1}{L} \sum_{l=1}^L C_{v|xgz}^l + (\mu_{v|xgz}^l - \frac{1}{L} \sum_{m=1}^L \mu_{v|xgz}^m)(\mu_{v|xgz}^l - \frac{1}{L} \sum_{m=1}^L \mu_{v|xgz}^m)^T \quad (71)$$

$$C_{v|x} \approx \text{E} [C_{v|xgz}^l] + \text{Cov} [\mu_{v|xgz}^l]_l \quad (72)$$

$$C_{v|xgz} = \left( \frac{z^2}{\sigma_x} A^T A + \left[ \sum_{k=1}^K g_k C_k \right]^{-1} \right)^{-1} \quad (73)$$

$$\mu_{v|xgz} = \frac{z}{\sigma_x} C_{v|xgz} A^T x \quad (74)$$

The mean of the finite mixture is given by

$$\mu_{v|x} \approx \frac{1}{L} \sum_{l=1}^L \mu_{v|xgz}^l \quad (75)$$

## 4 Appendix

### 4.1 Rules of differentiation

Assuming that  $y$  and  $a$  are vectors,  $S$  is a symmetric matrix of appropriate dimension,  $M$  is any matrix,  $f$  is a scalar function, and  $\gamma$  is a scalar variable.

$$\frac{\partial}{\partial y} y^T S y = 2S y \quad (76)$$

$$\frac{\partial}{\partial y} a^T y = a \quad (77)$$

$$\frac{\partial}{\partial S} y^T S^{-1} y = -S^{-1} y y^T S^{-1} \quad (78)$$

$$\frac{\partial}{\partial S} \log \det S = S^{-1} \quad (79)$$

$$\frac{\partial}{\partial \gamma} f(S(\gamma)) = \text{Tr} \left[ \frac{\partial f}{\partial S} \frac{\partial S}{\partial \gamma} \right] \quad (80)$$

$$\frac{\partial}{\partial \gamma} \gamma S = S \quad (81)$$

$$\frac{\partial}{\partial \gamma} f(\gamma) = f(\gamma) \frac{\partial}{\partial \gamma} \log f(\gamma) \quad (82)$$

$$\frac{\partial}{\partial S} \log \mathcal{N}(y; a, S) = -\frac{1}{2} [S^{-1} - S^{-1}(y - a)(y - a)^T S^{-1}] \quad (83)$$

$$\frac{\partial}{\partial a} (y - Ma)^T S (y - Ma) = -2M^T S (y - Ma) \quad (84)$$

### 4.2 Logarithms of used PDFs

$$\log \mathcal{N}(y; \mu, C) = -\frac{1}{2} [D \log(2\pi) + \log \det(C) + (y - \mu)^T C^{-1} (y - \mu)] \quad (85)$$

$$\log \text{Gam}(y; \alpha, \zeta) = \log(1) - \log(\Gamma(\alpha)) - \alpha \log(\zeta) + (\alpha - 1) \log(y) - \frac{y}{\zeta} \quad (86)$$

### 4.3 Switching and transforming variables of a Gaussian

We want to merge two Gaussians over  $x$  and  $v$  into one over  $v$

$$p(x | v, z) p(v | g) = \mathcal{N}(x; zAv, \sigma_x I) \mathcal{N}(v; 0, C_v) \quad (87)$$

The Gaussian over  $x$  spelled out is

$$\mathcal{N}(x; zAv, \sigma_x I) = \sqrt{\frac{1}{(2\pi)^{D_x} \sigma_x^{D_x}}} e^{-\frac{1}{2\sigma_x} (x - zAv)^T (x - zAv)} \quad (88)$$

rearranging the quadratic term:

$$\begin{aligned}
-\frac{1}{2}(x - zAv)^T(x - zAv) &= -\frac{1}{2}(x^T x - zv^T A^T x - zx^T Av + z^2 v^T A^T Av) = \\
&= -\frac{1}{2}(x - zAv)^T(x - zAv) = -\frac{1}{2}(x^T x - 2zx^T Av + z^2 v^T A^T Av) = \\
&= -\frac{x^T x}{2} + zx^T Av - \frac{z^2}{2} v^T A^T Av
\end{aligned} \tag{89}$$

as  $v^T A^T x = (x^T Av)^T$ , and both are scalars, thus equal to their transposes, it's also true that  $v^T A^T x = x^T Av$ . We have the identity for any symmetric matrix  $M$  and vector  $b$  that

$$-\frac{1}{2}v^T M v + b^T v = -\frac{1}{2}(v - M^{-1}b)^T M (v - M^{-1}b) + \frac{1}{2}b^T M^{-1}b \tag{90}$$

making the substitution  $M = z^2 A^T A$  and  $b = (zx^T A)^T = zA^T x$ , yielding  $M^{-1} = \frac{1}{z^2}(A^T A)^{-1}$  and  $M^{-1}b = \frac{1}{z}(A^T A)^{-1}A^T x = \frac{1}{z}A^+x$ , where  $A^+$  is the Moore-Penrose pseudoinverse of  $A$ . Thus we get

$$\begin{aligned}
&-\frac{1}{2}(x - zAv)^T(x - zAv) = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+x)^T z^2 A^T A (v - \frac{1}{z}A^+x) + \frac{1}{2}(A^T x)^T (A^T A)^{-1} A^T x = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+x)^T z^2 A^T A (v - \frac{1}{z}A^+x) + \frac{1}{2}x^T A A^{-1} A^{-T} A^T x = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+x)^T z^2 A^T A (v - \frac{1}{z}A^+x) + \frac{x^T x}{2} = \\
&= -\frac{1}{2}(v - \frac{1}{z}A^+x)^T z^2 A^T A (v - \frac{1}{z}A^+x)
\end{aligned} \tag{91}$$

which implies

$$e^{-\frac{1}{2\sigma_x}(x - zAv)^T(x - zAv)} = e^{-\frac{1}{2}(v - \frac{1}{z}A^+x)^T \frac{z^2}{\sigma_x} A^T A (v - \frac{1}{z}A^+x)} \tag{92}$$

meaning that

$$\mathcal{N}(x; zAv, \sigma_x I) = \alpha \mathcal{N}(v; \frac{1}{z}A^+x, \frac{\sigma_x}{z^2}(A^T A)^{-1}) \tag{93}$$

and as the formulas in the exponents are equal, the constant  $\alpha$  is given by the ratio of the normalisation terms

$$\sqrt{\frac{1}{(2\pi)^{D_x} \sigma_x^{D_x}}} = \alpha \sqrt{\frac{1}{(2\pi)^{D_v} \det(\frac{\sigma_x}{z^2} (A^T A)^{-1})}} \quad (94)$$

$$\alpha = \sqrt{\frac{(2\pi)^{D_v} \frac{\sigma_x^{D_v}}{z^{2D_v}} \det((A^T A)^{-1})}{(2\pi)^{D_x} \sigma_x^{D_x}}} \quad (95)$$

$$\alpha = \sqrt{\frac{(2\pi)^{D_v} \sigma_x^{D_v}}{(2\pi)^{D_x} \sigma_x^{D_x} z^{2D_v} \det(A^T A)}} \quad (96)$$

making the simplifying assumption  $D_x = D_v$  we arrive to

$$\mathcal{N}(x; zAv, \sigma_x I) = \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(v; \frac{1}{z} A^+ x, \frac{\sigma_x}{z^2} (A^T A)^{-1}) \quad (97)$$

#### 4.4 Merging two Gaussian distributions

$$\mathcal{N}(v; \mu_1, C_1) \mathcal{N}(v; \mu_2, C_2) = \mathcal{N}(\mu_1; \mu_2, C_1 + C_2) \mathcal{N}(v; \mu_c, C_c) \quad (98)$$

where  $C_c = (C_1^{-1} + C_2^{-1})^{-1}$  and  $\mu_c = C_c(C_1^{-1}\mu_1 + C_2^{-1}\mu_2)$ . Substitution to these formulas yields

$$\begin{aligned} & \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(v; \frac{1}{z} A^+ x, \frac{\sigma_x}{z^2} (A^T A)^{-1}) \mathcal{N}(v; 0, C_v) = \\ & \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(\frac{1}{z} A^+ x; 0, \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v) \mathcal{N}(v; \mu_c, C_c) \end{aligned} \quad (99)$$

$$C_c = (\frac{z^2}{\sigma_x} (A^T A) + C_v^{-1})^{-1} \quad (100)$$

$$\mu_c = C_c \frac{z}{\sigma_x} (A^T A) A^+ x = \frac{z}{\sigma_x} C_c A^T x \quad (101)$$

#### 4.5 Equivalence of the two likelihood formulas of CSM

Expanding [10](#) yields

$$\begin{aligned} p(x | z, g) &= \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \frac{1}{\sqrt{(2\pi)^{D_v} \det C(z, g)}} e^{-\frac{1}{2} (\frac{1}{z} A^+ x)^T C^{-1}(z, g) \frac{1}{z} A^+ x} = \\ &= \frac{1}{\sqrt{(2\pi z^2)^{D_v} \det(A^T A C(z, g))}} e^{-\frac{1}{2z^2} x^T A^{+T} C^{-1}(z, g) A^+ x} \quad (102) \\ & \quad C(z, g) \equiv \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \quad (103) \end{aligned}$$

so in the exponent, in the place of the covariance matrix, we have

$$\begin{aligned}
& \frac{1}{z^2} ((A^T A)^{-1} A^T)^T \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right]^{-1} (A^T A)^{-1} A^T = \\
& = \frac{1}{z^2} A (A^T A)^{-1} \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right]^{-1} (A^T A)^{-1} A^T = \\
& = \frac{1}{z^2} A \left[ \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right] (A^T A) \right]^{-1} (A^T A)^{-1} A^T = \\
& = \frac{1}{z^2} A \left[ \frac{\sigma_x}{z^2} I + C_v A^T A \right]^{-1} (A^T A)^{-1} A^T = \\
& = \frac{1}{z^2} A \left[ (A^T A) \left[ \frac{\sigma_x}{z^2} I + C_v A^T A \right] \right]^{-1} A^T = \\
& = \frac{1}{z^2} A \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right]^{-1} A^T
\end{aligned} \tag{104}$$

assuming that  $A$  is invertible this is equal to

$$\begin{aligned}
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right]^{-1} [A^{-T}]^{-1} = \\
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ A^{-T} \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right] \right]^{-1} = \\
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ \frac{\sigma_x}{z^2} A + A C_v A^T A \right]^{-1} = \\
& \frac{1}{z^2} \left[ \left[ \frac{\sigma_x}{z^2} A + A C_v A^T A \right] A^{-1} \right]^{-1} = \\
& \frac{1}{z^2} \left[ \frac{\sigma_x}{z^2} I + A C_v A^T \right]^{-1} = \\
& = [\sigma_x I + z^2 A C_v A^T]^{-1}
\end{aligned} \tag{105}$$

under the square root we have

$$\begin{aligned}
& z^{2D_v} \det(A^T A C(z, g)) = z^{2D_v} \det(A^T) \det(A) \det(C(z, g)) = z^{2D_v} \det(A) \det(C(z, g)) \det(A^T) \\
& = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v\right) \det(A^T) = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v\right) \det(A^T) = \\
& = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v\right) \det(A^T) = z^{2D_v} \det \left[ A \left[ \frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v \right] A^T \right] = \\
& = z^{2D_v} \det \left[ \frac{\sigma_x}{z^2} I + A C_v A^T \right] = \det [\sigma_x I + z^2 A C_v A^T]
\end{aligned} \tag{106}$$

## 4.6 Precision component formulation of the CSM model

The model can be equally well parametrised by precision components

$$p(v | g) = \mathcal{N}(v; 0, \Lambda_v^{-1}) \quad (107)$$

$$\Lambda_v = \sum_{k=1}^K g_k \Lambda_k \quad (108)$$

in this case the conditional posterior over  $v$  takes the form

$$p(v | x, g) = \mathcal{N}\left(v; \frac{1}{\sigma_x} \left(\frac{1}{\sigma_x} A^T A + \Lambda_v\right)^{-1} A^T x, \left(\frac{1}{\sigma_x} A^T A + \Lambda_v\right)^{-1}\right) \quad (109)$$

and the conditional posterior of  $g$  will look as follows

$$\log p(g | X, V) \sim -\frac{1}{2} [\log(\det(\Lambda_v^{-1})) + v^T \Lambda_v v] + \log p(g) \quad (110)$$

The gradient of the expectation of the complete-data log-likelihood with respect to the joint posterior will look like this

$$\Lambda_k = U_k^T U_k \quad (111)$$

$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \frac{1}{L} \sum_{m=1}^{LN} g_k^m \text{Tr} \left[ \left[ (\Lambda_v^m)^{-1} - v^m v^{mT} \right] \hat{U}_k^{ij} \right] \quad (112)$$

## 4.7 Batches of observations

For a single set of component activations  $g$  and contrast  $z$ , we might have a batch of  $v$  and  $x$  values of size  $B$ . This modifies expressions as follows.

Conditional posterior of  $g$  (Eq. 58)

$$\log p(g | X, V) \sim -\frac{1}{2} \left[ B \log(\det(C_v)) + \sum_{b=1}^B v_b^T C_v^{-1} v_b \right] + \log p(g) \quad (113)$$

Conditional posterior of  $z$  (Eq. 61)

$$\log p(z | X, V) \sim -\frac{1}{2} \left[ B D_x \log(\sigma_x) + \frac{1}{\sigma_x} \sum_{b=1}^B (x_b - z A v_b)^T (x_b - z A v_b) \right] + \log p(z) \quad (114)$$

Matrix formula in the derivative of the expectation of the complete-data log-likelihood (Eq. 40)

$$M_k = -\frac{1}{2L} \left[ \sum_{m=1}^{LN} g_k^m \left[ B (C_v^m)^{-1} - (C_v^m)^{-1} \left( \sum_{b=1}^B v^{m,b} v^{(m,b)T} \right) (C_v^m)^{-1} \right] \right] \quad (115)$$

## 4.8 Latents affecting the mean

of  $v$ , instead of the covariance

$$p_m(v \mid g) = \mathcal{N}(v; Bg, \sigma_v I) \quad (116)$$

$$p_m(v \mid x, g, z) = \mathcal{N}(v; \mu_m, C_m) \quad (117)$$

$$C_m = \left( \frac{z^2}{\sigma_x} A^T A + \frac{1}{\sigma_v} I \right)^{-1} \quad (118)$$

$$\mu_m = C_m \left( \frac{z}{\sigma_x} A^T x + \frac{1}{\sigma_v} Bg \right) \quad (119)$$

likelihood of  $g$  and  $z$ , by intuition:

$$p_m(x \mid g, z) = \mathcal{N}(x; zABg, \sigma_x I + z^2 \sigma_v AA^T) \quad (120)$$

log-posterior of  $g$  and  $z$

$$\begin{aligned} \log p_m(g, z \mid x) \sim & -\frac{1}{2} [\log \det(C_{xm}) + (x - zABg)^T C_{xm}^{-1} (x - zABg)] + \\ & + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} + (\alpha_g - 1) \sum_{k=1}^K \log(g_k) - \frac{1}{\theta_g} \sum_{k=1}^K g_k \end{aligned} \quad (121)$$

$$C_{xm} = \sigma_x I + z^2 \sigma_v AA^T \quad (122)$$

the derivative w.r.t.  $g$  is given by Eq. 84

$$\frac{\partial}{\partial g} \log p_m(g, z \mid x) = zB^T A^T C_{xm}^{-1} (x - zAB) + \frac{\partial}{\partial g} \log p(g) \quad (123)$$

for the sake of the derivative w.r.t.  $z$ , we need write out the quadratic form:

$$\begin{aligned} \mathcal{L}_m &\equiv (x - zABg)^T C_{xm}^{-1} (x - zABg) \equiv (x - zy)^T C_{xm}^{-1} (x - zy) \\ &= x^T C_{xm}^{-1} x - 2x^T C_{xm}^{-1} zy + zy^T C_{xm}^{-1} zy = \\ &= x^T C_{xm}^{-1} x - 2x^T \left[ \frac{1}{z} C_{xm} \right]^{-1} y + y^T \left[ \frac{1}{z^2} C_{xm} \right]^{-1} y \end{aligned} \quad (124)$$

$$\begin{aligned} \frac{\partial \mathcal{L}_m}{\partial z} = & \text{Tr} \left[ \frac{\partial}{\partial C_{xm}} x^T C_{xm}^{-1} x \frac{\partial C_{xm}}{\partial z} \right] + \text{Tr} \left[ \frac{\partial}{\partial \frac{1}{z} C_{xm}} 2x^T \left[ \frac{1}{z} C_{xm} \right]^{-1} y \frac{\partial \frac{1}{z} C_{xm}}{\partial z} \right] + \\ & + \text{Tr} \left[ \frac{\partial}{\partial \frac{1}{z^2} C_{xm}} y^T \left[ \frac{1}{z^2} C_{xm} \right]^{-1} y \frac{\partial \frac{1}{z^2} C_{xm}}{\partial z} \right] \end{aligned} \quad (125)$$

the six derivatives being



$$\frac{\partial}{\partial C_{xm}} x^T C_{xm}^{-1} x = -C_{xm}^{-1} x x^T C_{xm}^{-1} \quad (126)$$

$$\frac{\partial}{\partial \frac{1}{z} C_{xm}} 2x^T \left[ \frac{1}{z} C_{xm} \right]^{-1} y = -2z^2 C_{xm}^{-1} x y^T C_{xm}^{-1} \quad (127)$$

$$\frac{\partial}{\partial \frac{1}{z^2} C_{xm}} y^T \left[ \frac{1}{z^2} C_{xm} \right]^{-1} y = -z^4 C_{xm}^{-1} y y^T C_{xm}^{-1} \quad (128)$$

$$\frac{\partial C_{xm}}{\partial z} = 2z \sigma_v A A^T \quad (129)$$

$$\frac{\partial \frac{1}{z} C_{xm}}{\partial z} = -\frac{\sigma_x}{z^2} I + \sigma_v A A^T \quad (130)$$

$$\frac{\partial \frac{1}{z^2} C_{xm}}{\partial z} = -\frac{2\sigma_x}{z^3} I \quad (131)$$

substituting back

$$\begin{aligned} \frac{\partial \mathcal{L}_m}{\partial z} &= \text{Tr} \left[ -C_{xm}^{-1} \left[ 2z \sigma_v A A^T x x^T - 2\sigma_x x y^T + 2z^2 \sigma_v A A^T x y^T - 2z \sigma_x y y^T \right] C_{xm}^{-1} \right] = \\ &= -2 \text{Tr} \left[ C_{xm}^{-1} \left[ z \sigma_v A A^T (x x^T + z x y^T) - \sigma_x (x y^T - z y y^T) \right] C_{xm}^{-1} \right] \end{aligned} \quad (132)$$