

# 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. 3.5):

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

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

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

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

## 2.2 Log-likelihood of the parameters

$$p(X | U_{1..K}) = \prod_{n=1}^N p(x_n | U_{1..K}) \quad (11)$$

$$p(X | U_{1..K}) = \prod_{n=1}^N \int \int_{-\infty}^{\infty} p(x_n | z, g) p(g) p(z) dg dz \quad (12)$$

$$(13)$$

approximation of the integrals by samples from the priors  $p(g_n)$  and  $p(z_n)$

$$p(X | U_{1..K}) \approx \prod_{n=1}^N \frac{1}{L} \sum_{l=1}^L p(x_n | z^l, g^l) \quad (14)$$

using [10](#)

$$p(X | U_{1..K}) \approx \left( L \sqrt{\det(A^T A)} \right)^{-N} \prod_{n=1}^N \sum_{l=1}^L h(z^l) \mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l)) \quad (15)$$

$$\log p(X | U_{1..K}) \approx -N(\log L + \frac{1}{2} \log(\det(A^T A))) + \sum_{n=1}^N \log \left[ \sum_{l=1}^L h(z^l) \mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l)) \right] \quad (16)$$

$$h^l \equiv h(z^l), \quad f_n^l \equiv f(x_n, z^l), \quad C^l \equiv C(z^l, g^l) \quad (17)$$

$$\mathcal{L}_n^l \equiv h^l \mathcal{N}(f_n^l; 0, C^l), \quad \mathcal{L}_n \equiv \sum_{l=1}^L \mathcal{L}_n^l \quad (18)$$

$$\log p(X | U_{1..K}) \approx -N(\log L + \frac{1}{2} \log(\det(A^T A))) + \sum_{n=1}^N \log \mathcal{L}_n \quad (19)$$

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

$$\begin{aligned} \frac{\partial \log p(X | U_{1..K})}{\partial [U_k]_{i,j}} &\approx \sum_{n=1}^N \frac{\partial \log \mathcal{L}_n}{\partial [U_k]_{i,j}} = \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \frac{\partial \mathcal{L}_n}{\partial [U_k]_{i,j}} = \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L \frac{\partial \mathcal{L}_n^l}{\partial [U_k]_{i,j}} = \\ &= \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L \text{Tr} \left[ \frac{\partial \mathcal{L}_n^l}{\partial C^l} \frac{\partial C^l}{\partial [U_k]_{i,j}} \right] \end{aligned} \quad (20)$$

the derivatives in this formula are the following

$$\begin{aligned} \frac{\partial \mathcal{L}_n^l}{\partial C^l} &= h^l \frac{\partial}{\partial C^l} \mathcal{N}(f_n^l; 0, C^l) = h^l \mathcal{N}(f_n^l; 0, C^l) \frac{\partial}{\partial C^l} \log \mathcal{N}(f_n^l; 0, C^l) = \\ &= -\frac{h^l}{2} \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \end{aligned} \quad (21)$$

$$\frac{\partial C(z, g)}{\partial [U_k]_{i,j}} = g_k \frac{\partial (U_k^T U_k)}{\partial [U_k]_{i,j}} = g_k (U_k^T J^{ij} + J^{ji} U_k) \equiv g_k \hat{U}_k^{ij} \quad (22)$$

substituting back to the derivative

$$\begin{aligned} \frac{\partial \log p(X | U_{1..K})}{\partial [U_k]_{i,j}} &\approx \\ &\approx -\frac{1}{2} \sum_{n=1}^N \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) \text{Tr} \left[ [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \hat{U}_k^{ij} \right] = \\ &= -\frac{1}{2} \text{Tr} \left[ \sum_{n=1}^N \left( \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \right) \hat{U}_k^{ij} \right] \end{aligned} \quad (23)$$

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

$$M_k = -\frac{1}{2} \sum_{n=1}^N \left( \frac{1}{\mathcal{L}_n} \sum_{l=1}^L h^l g_k^l \mathcal{N}(f_n^l; 0, C^l) [(C^l)^{-1} - (C^l)^{-1} f_n^l (f_n^l)^T (C^l)^{-1}] \right) \quad (24)$$

$$\frac{\partial \log p(X | U_{1..K})}{\partial [U_k]_{i,j}} \approx \text{Tr} [M_k \hat{U}_k^{ij}] = \sum_{a=1}^{Dv} [M_k]_{j,a} [U_k]_{i,a} + [M_k]_{a,j} [U_k]_{i,a} \quad (25)$$

## 2.3 Complete-data log-likelihood

$$p(V, G, Z, X | U_{1..K}) = \prod_{n=1}^N p(x_n | v_n, z_n) p(v_n | g_n) p(g_n) p(z_n) \quad (26)$$

the logarithm of this will be

$$\begin{aligned}
\log p(V, G, Z, X \mid U_{1..K}) &= \\
&= \sum_{n=1}^N [\log p(g_n) + \log p(z_n) + \log p(x_n \mid v_n) + \log p(v_n \mid g_n)] = \\
&= N(\log p(g) + \log p(z)) + \sum_{n=1}^N \log p(x_n \mid v_n) + \log p(v_n \mid g_n) = \\
&= c + \sum_{n=1}^N \log p(v_n \mid g_n)
\end{aligned} \tag{27}$$

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

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

$$\mathcal{L} = \int \int_{-\infty}^{\infty} p(V, G, Z \mid X) \log p(V, G, Z, X \mid U_{1..K}) dV dG dZ. \tag{28}$$

We can approximate this integral by averaging over  $L$  samples from the full posterior, separately for each observation  $x_n$ , as defined in Seq. ?? . 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  $p(v \mid g)$ . This way we arrive to the following expression

$$\begin{aligned}
\mathcal{L} &\sim \sum_{n=1}^N \frac{1}{L} \sum_{l=1}^L -\frac{1}{2} \left[ \log \left( \det \left( C_v^{(l,n)} \right) \right) + v^{(l,n)T} \left( C_v^{l,n} \right)^{-1} v^{l,n} \right] = \\
&= -\frac{1}{2L} \sum_{m=1}^{NL} \left[ \log \left( \det \left( C_v^m \right) \right) + v^{mT} \left( C_v^m \right)^{-1} v^m \right]
\end{aligned} \tag{29}$$

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. ?? , by the chain rule, the derivative of  $\mathcal{L}$  according to an element of  $U_k$  looks like this

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} &= -\frac{1}{2L} \sum_{m=1}^{NL} \text{Tr} \left[ \frac{\partial \mathcal{L}^m}{\partial C_v^m} \frac{\partial C_v^m}{\partial [U_k]_{i,j}} \right] = \\
&= -\frac{1}{2L} \sum_{m=1}^{LN} \text{Tr} \left[ \left[ (C_v^m)^{-1} - (C_v^m)^{-1} v^m v^{mT} (C_v^m)^{-1} \right] g_k^m \hat{U}_k^{ij} \right] = \\
&= -\frac{1}{2L} \text{Tr} \left[ \sum_{m=1}^{LN} g_k^m \left[ (C_v^m)^{-1} - (C_v^m)^{-1} (v^m v^{mT}) (C_v^m)^{-1} \right] \hat{U}_k^{ij} \right]
\end{aligned} \tag{30}$$

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

$$M_k = -\frac{1}{2L} \left[ \sum_{m=1}^{LN} g_k^m \left[ (C_v^m)^{-1} - (C_v^m)^{-1} (v^m v^{mT}) (C_v^m)^{-1} \right] \right] \tag{31}$$

$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \sum_{a=1}^{Dv} [M_K]_{j,a} [U_k]_{i,a} + [M_k]_{a,j} [U_k]_{i,a} \tag{32}$$

### 3 Posteriors

#### 3.1 Full posterior

$$p(v, g, z | x) = p(x | v, g, z) p(v | g) p(g) p(z) \frac{1}{p(x)} \sim p(x | v, z) p(v | g) p(g) p(z) \tag{33}$$

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:

$$\begin{aligned}
\log p(v, g, z | x) &\sim \log p(x | v, z) + \log p(v | g) + \log p(g) + \log p(z) = \\
&= \log \mathcal{N}(x; zAv, \sigma_x I) + \log \mathcal{N}(v; 0, C_v) + \log \text{Gam}(g; \alpha_g, \theta_g) + \log \text{Gam}(z; \alpha_z, \theta_z)
\end{aligned} \tag{34}$$

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

$$\begin{aligned}
\log p(v, g, z | x) &\sim -\frac{1}{2\sigma_x} (x - zAv)^T (x - zAv) - \\
&\quad -\frac{1}{2} \left[ \log(\det(C_v)) + v^T C_v^{-1} v \right] + \\
&\quad + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z}
\end{aligned} \tag{35}$$

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

$$\begin{aligned} \log p(v, g, z | x) \sim & -\frac{z}{2\sigma_x} (zv^T A^T Av - 2x^T Av) - \\ & -\frac{1}{2} [\log(\det(C_v)) + v^T C_v^{-1} v] + \\ & + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + \\ & + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} \end{aligned} \quad (36)$$

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

$$\log p(v, g, z | x) \sim -\frac{1}{2} \left[ \frac{z^2}{\sigma_x} v^T A^T Av - \frac{2z}{\sigma_x} x^T Av + v^T C_v^{-1} v \right] + f_1(g, z) \quad (37)$$

lumping the two quadratic forms together

$$\log p(v, g, z | x) \sim \frac{z}{\sigma_x} x^T Av - \frac{1}{2} v^T \left[ \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right] v + f_1(g, z) \quad (38)$$

Taking the derivative using Eq. 32 and 33 we get

$$\frac{\partial}{\partial v} \log p(v, g, z | x) = \frac{z}{\sigma_x} A^T x - \left[ \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right] v \quad (39)$$

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

$$\begin{aligned} \log p(v, g, z | x) \sim & -\frac{1}{2} [\log \det(C_v) + v^T C_v^{-1} v] + \\ & + \sum_{j=1}^K \left[ (\alpha_g - 1) \log(g_j) - \frac{g_j}{\theta_g} \right] + f_2(v, z) \end{aligned} \quad (40)$$

Taking the derivative w.r.t. a single  $g_i$  using Eq. 36 we get

$$\begin{aligned} \frac{\partial}{\partial g_i} \log p(v, g, z | x) = & -\frac{1}{2} \text{Tr} \left[ \frac{\partial}{\partial C_v} [\log \det(C_v) + v^T C_v^{-1} v] \frac{\partial C_v}{\partial g_i} \right] + \\ & + \frac{\partial}{\partial g_i} \left[ (\alpha_g - 1) \log(g_i) - \frac{g_i}{\theta_g} \right] \end{aligned} \quad (41)$$

using Eq. 35, 34 and 37 we arrive to

$$\frac{\partial}{\partial g_i} \log p(v, g, z | x) = -\frac{1}{2} \text{Tr} \left[ [C_v^{-1} - C_v^{-1} v v^T C_v^{-1}] C_i \right] + \frac{\alpha_g - 1}{g_i} - \frac{1}{\theta_g} \quad (42)$$

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

$$\log p(v, g, z | x) \sim -\frac{z}{2\sigma_x} (zv^T A^T Av - 2x^T Av) + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} + f_3(g, v) \quad (43)$$

$$\frac{\partial}{\partial z} \log p(v, g, z | x) = \frac{1}{\sigma_x} [x^T Av - zv^T A^T Av] + \frac{\alpha_z - 1}{z} - \frac{1}{\theta_z} \quad (44)$$

## 3.2 Conditional posteriors

### 3.2.1 Conditional posterior of $v$

$$p(v | x, g, z) = \frac{p(x | v, z, g)p(v | z, g)}{p(x | z, g)} = \frac{\mathcal{N}(x; zAv, \sigma_x I) \mathcal{N}(v; 0, C_v)}{\int_{-\infty}^{\infty} \mathcal{N}(x; zAv, \sigma_x I) \mathcal{N}(v; 0, C_v) dv} \quad (45)$$

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

$$\mathcal{N}(x; zAv, \sigma_x I) \mathcal{N}(v; 0, C_v) = c \mathcal{N}(v; \mu_{post}, C_{post}) \quad (46)$$

The denominator of Eq. ?? 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

$$p(v | x, g, z) = \mathcal{N} \left( v; \frac{z}{\sigma_x} \left( \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right)^{-1} A^T x, \left( \frac{z^2}{\sigma_x} A^T A + C_v^{-1} \right)^{-1} \right) \quad (47)$$

### 3.2.2 Conditional posterior of $g$

$$p(g | X, V, z) = \frac{p(X | g, V, z)p(g | V, z)}{p(X | V, z)} = \frac{p(V | g)p(g)}{p(V)} \quad (48)$$

taking the logarithm and discarding constant terms

$$\log p(g | X, V) \sim -\frac{1}{2} [\log(\det(C_v)) + v^T C_v^{-1} v] + \log p(g) \quad (49)$$

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

$$\begin{aligned} \log p(g_j | g_{\neg j}, X, V) &= \frac{p(V | g_j, g_{\neg j}, X)p(g_j | g_{\neg j}, X)}{p(V | g_{\neg j}, X)} = \\ &= \frac{p(V | g)p(g_j)}{p(V | g_{\neg j})} \sim p(V | g)p(g_j) \end{aligned} \quad (50)$$

### 3.2.3 Conditional posterior of $z$

$$p(z \mid X, V, g) = \frac{p(X \mid g, z, V)p(z \mid V, g)}{p(X \mid V, g)} \sim p(X \mid z, V)p(z) \quad (51)$$

the log-posterior being

$$\log p(z \mid X, V) \sim -\frac{1}{2} \left[ D_x \log(\sigma_x) + \frac{1}{\sigma_x} (x - zAv)^T (x - zAv) \right] + \log p(z) \quad (52)$$

## 3.3 Marginal posteriors

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

$$p(g, z \mid x) \sim p(x \mid g, z)p(g)p(z) \quad (53)$$

from Eq. 6

$$\log p(g, z \mid x) \sim -\frac{1}{2} [\log \det(C_x) + x^T C_x^{-1} x] + (\alpha_z - 1) \log(z) - \frac{z}{\theta_z} + (\alpha_g - 1) \sum_{k=1}^K \log(g_k) - \frac{1}{\theta_g} \sum_{k=1}^K g_k \quad (54)$$

$$C_x = \sigma_x I + z^2 A \left( \sum_{k=1}^K g_k U_k^T U_k \right) A^T \quad (55)$$

### 3.3.2 Marginal posterior and MAP of $g$

$$g_{MAP} = \arg \max_g p(g \mid x) = \arg \max_g \frac{p(x \mid g)p(g)}{p(x)} = \arg \max_g p(x \mid g)p(g) \quad (56)$$

$$p(x \mid g) = \int_{-\infty}^{\infty} p(x \mid z, g)p(z)dz \approx \frac{1}{L} \sum_{l=1}^L p(x \mid g, z^l) \quad (57)$$

## 4 Appendix

### 4.1 Rules of differentiation

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



$$\frac{\partial}{\partial y} y^T M y = 2M y \quad (58)$$

$$\frac{\partial}{\partial y} a^T y = a \quad (59)$$

$$\frac{\partial}{\partial M} y^T M^{-1} y = -M^{-1} y y^T M^{-1} \quad (60)$$

$$\frac{\partial}{\partial M} \log \det M = M^{-1} \quad (61)$$

$$\frac{\partial}{\partial s} f(M(s)) = \text{Tr} \left[ \frac{\partial f}{\partial M} \frac{\partial M}{\partial s} \right] \quad (62)$$

$$\frac{\partial}{\partial s} s M = M \quad (63)$$

$$\frac{\partial}{\partial s} f(s) = f(s) \frac{\partial}{\partial s} \log f(s) \quad (64)$$

$$\frac{\partial}{\partial M} \log \mathcal{N}(y; a, M) = M^{-1} - M^{-1} (y - a)(y - a)^T M^{-1} \quad (65)$$

## 4.2 Logarithms of used PDFs

$$\log \mathcal{N}(y; \mu, C) = -\frac{1}{2} [D \log(2\pi) + \log \det(C) + (y - \mu)^T C^{-1} (y - \mu)] \quad (66)$$

$$\log \text{Gam}(y; \alpha, \theta) = \log(1) - \log(\Gamma(\alpha)) - \alpha \log(\theta) + (\alpha - 1) \log(y) - \frac{y}{\theta} \quad (67)$$

## 4.3 Merging two Gaussian distributions

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

$$p(x | v, z) p(v | g) = \mathcal{N}(x; zAv, \sigma_x I) \mathcal{N}(v; 0, C_v) \quad (68)$$

The Gaussian over  $x$  spelled out is

$$\mathcal{N}(x; zAv, \sigma_x I) = \sqrt{\frac{1}{(2\pi)^{D_x} \sigma_x^{D_x}}} e^{-\frac{1}{2\sigma_x} (x - zAv)^T (x - zAv)} \quad (69)$$

rearranging the quadratic term:

$$\begin{aligned} -\frac{1}{2} (x - zAv)^T (x - zAv) &= -\frac{1}{2} (x^T x - z v^T A^T x - z x^T A v + z^2 v^T A^T A v) = \\ &= -\frac{1}{2} (x - zAv)^T (x - zAv) = -\frac{1}{2} (x^T x - 2z x^T A v + z^2 v^T A^T A v) = \\ &= -\frac{x^T x}{2} + z x^T A v - \frac{z^2}{2} v^T A^T A v \end{aligned} \quad (70)$$

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

$$-\frac{1}{2}v^T M v + b^T v = -\frac{1}{2}(v - M^{-1}b)^T M (v - M^{-1}b) + \frac{1}{2}b^T M^{-1}b \quad (71)$$

making the substitution  $M = z^2 A^T A$  and  $b = (z x^T A)^T = z A^T x$ , 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

$$\begin{aligned} & -\frac{1}{2}(x - z A v)^T (x - z A v) = \\ & = -\frac{x^T x}{2} - \frac{1}{2}\left(v - \frac{1}{z}A^+ x\right)^T z^2 A^T A \left(v - \frac{1}{z}A^+ x\right) + \frac{1}{2}(A^T x)^T (A^T A)^{-1} A^T x = \\ & = -\frac{x^T x}{2} - \frac{1}{2}\left(v - \frac{1}{z}A^+ x\right)^T z^2 A^T A \left(v - \frac{1}{z}A^+ x\right) + \frac{1}{2}x^T A A^{-1} A^{-T} A^T x = \\ & = -\frac{x^T x}{2} - \frac{1}{2}\left(v - \frac{1}{z}A^+ x\right)^T z^2 A^T A \left(v - \frac{1}{z}A^+ x\right) + \frac{x^T x}{2} = \\ & = -\frac{1}{2}\left(v - \frac{1}{z}A^+ x\right)^T z^2 A^T A \left(v - \frac{1}{z}A^+ x\right) \end{aligned} \quad (72)$$

which implies

$$e^{-\frac{1}{2\sigma_x}(x - z A v)^T (x - z A v)} = e^{-\frac{1}{2}\left(v - \frac{1}{z}A^+ x\right)^T \frac{z^2}{\sigma_x} A^T A \left(v - \frac{1}{z}A^+ x\right)} \quad (73)$$

meaning that

$$\mathcal{N}(x; z A v, \sigma_x I) = \alpha \mathcal{N}\left(v; \frac{1}{z}A^+ x, \frac{\sigma_x}{z^2}(A^T A)^{-1}\right) \quad (74)$$

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

$$\sqrt{\frac{1}{(2\pi)^{D_x} \sigma_x^{D_x}}} = \alpha \sqrt{\frac{1}{(2\pi)^{D_v} \det(\frac{\sigma_x}{z^2}(A^T A)^{-1})}} \quad (75)$$

$$\alpha = \sqrt{\frac{(2\pi)^{D_v} \frac{\sigma_x^{D_v}}{z^{2D_v}} \det((A^T A)^{-1})}{(2\pi)^{D_x} \sigma_x^{D_x}}} \quad (76)$$

$$\alpha = \sqrt{\frac{(2\pi)^{D_v} \sigma_x^{D_v}}{(2\pi)^{D_x} \sigma_x^{D_x} z^{2D_v} \det(A^T A)}} \quad (77)$$

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

$$\mathcal{N}(x; zAv, \sigma_x I) = \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(v; \frac{1}{z} A^+ x, \frac{\sigma_x}{z^2} (A^T A)^{-1}) \quad (78)$$

we can merge two Gaussian distributions over  $v$  into one by using the following formula

$$\mathcal{N}(v; \mu_1, C_1) \mathcal{N}(v; \mu_2, C_2) = \mathcal{N}(\mu_1; \mu_2, C_1 + C_2) \mathcal{N}(v; \mu_c, C_c) \quad (79)$$

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

$$\begin{aligned} & \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(v; \frac{1}{z} A^+ x, \frac{\sigma_x}{z^2} (A^T A)^{-1}) \mathcal{N}(v; 0, C_v) = \\ & \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(\frac{1}{z} A^+ x; 0, \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v) \mathcal{N}(v; \mu_c, C_c) \end{aligned} \quad (80)$$

$$C_c = (\frac{z^2}{\sigma_x} (A^T A) + C_v^{-1})^{-1} \quad (81)$$

$$\mu_c = C_c \frac{z}{\sigma_x} (A^T A) A^+ x = \frac{z}{\sigma_x} C_c A^T x \quad (82)$$

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

Expanding 10 yields

$$\begin{aligned} p(x | z, g) &= \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \frac{1}{\sqrt{(2\pi)^{D_v} \det C(z, g)}} e^{-\frac{1}{2} (\frac{1}{z} A^+ x)^T C^{-1}(z, g) \frac{1}{z} A^+ x} = \\ &= \frac{1}{\sqrt{(2\pi z^2)^{D_v} \det(A^T A C(z, g))}} e^{-\frac{1}{2z^2} x^T A^{+T} C^{-1}(z, g) A^+ x} \quad (83) \\ & \quad C(z, g) \equiv \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \quad (84) \end{aligned}$$

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

$$\begin{aligned} & \frac{1}{z^2} ((A^T A)^{-1} A^T)^T \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right]^{-1} (A^T A)^{-1} A^T = \\ &= \frac{1}{z^2} A (A^T A)^{-1} \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right]^{-1} (A^T A)^{-1} A^T = \\ &= \frac{1}{z^2} A \left[ \left[ \frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v \right] (A^T A) \right]^{-1} (A^T A)^{-1} A^T = \\ &= \frac{1}{z^2} A \left[ \frac{\sigma_x}{z^2} I + C_v A^T A \right]^{-1} (A^T A)^{-1} A^T = \\ &= \frac{1}{z^2} A \left[ (A^T A) \left[ \frac{\sigma_x}{z^2} I + C_v A^T A \right] \right]^{-1} A^T = \\ &= \frac{1}{z^2} A \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right]^{-1} A^T \end{aligned} \quad (85)$$

assuming that  $A$  is invertible this is equal to

$$\begin{aligned}
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right]^{-1} [A^{-T}]^{-1} = \\
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ A^{-T} \left[ \frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A \right] \right]^{-1} = \\
& \frac{1}{z^2} [A^{-1}]^{-1} \left[ \frac{\sigma_x}{z^2} A + A C_v A^T A \right]^{-1} = \\
& \frac{1}{z^2} \left[ \left[ \frac{\sigma_x}{z^2} A + A C_v A^T A \right] A^{-1} \right]^{-1} = \\
& \frac{1}{z^2} \left[ \frac{\sigma_x}{z^2} I + A C_v A^T \right]^{-1} = \\
& = [\sigma_x I + z^2 A C_v A^T]^{-1}
\end{aligned} \tag{86}$$

under the square root we have

$$\begin{aligned}
& z^{2D_v} \det(A^T A C(z, g)) = z^{2D_v} \det(A^T) \det(A) \det(C(z, g)) = z^{2D_v} \det(A) \det(C(z, g)) \det(A^T) \\
& = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v\right) \det(A^T) = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} (A^T A)^{-1} + C_v\right) \det(A^T) = \\
& = z^{2D_v} \det(A) \det\left(\frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v\right) \det(A^T) = z^{2D_v} \det \left[ A \left[ \frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v \right] A^T \right] = \\
& = z^{2D_v} \det \left[ \frac{\sigma_x}{z^2} I + A C_v A^T \right] = \det [\sigma_x I + z^2 A C_v A^T]
\end{aligned} \tag{87}$$

#### 4.5 Precision component formulation of the CSM model

The model can be equally well parametrised by precision components

$$p(v | g) = \mathcal{N}(v; 0, \Lambda_v^{-1}) \tag{88}$$

$$\Lambda_v = \sum_{k=1}^K g_k \Lambda_k \tag{89}$$

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

$$p(v | x, g) = \mathcal{N} \left( v; \frac{1}{\sigma_x} \left( \frac{1}{\sigma_x} A^T A + \Lambda_v \right)^{-1} A^T x, \left( \frac{1}{\sigma_x} A^T A + \Lambda_v \right)^{-1} \right) \tag{90}$$

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

$$\log p(g | X, V) \sim -\frac{1}{2} [\log(\det(\Lambda_v^{-1})) + v^T \Lambda_v v] + \log p(g) \tag{91}$$

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

$$\Lambda_k = U_k^T U_k \quad (92)$$

$$\frac{\partial \mathcal{L}}{\partial [U_k]_{i,j}} = \frac{1}{L} \sum_{m=1}^{LN} g_k^m \text{Tr} \left[ \left[ (\Lambda_v^m)^{-1} - v^m v^{mT} \right] \hat{U}_k^{ij} \right] \quad (93)$$

#### 4.6 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. ??)

$$\log p(g \mid X, V) \sim -\frac{1}{2} \left[ B \log(\det(C_v)) + \sum_{b=1}^B v_b^T C_v^{-1} v_b \right] + \log p(g) \quad (94)$$

Conditional posterior of  $z$  (Eq. ??)

$$\log p(z \mid X, V) \sim -\frac{1}{2} \left[ B D_x \log(\sigma_x) + \frac{1}{\sigma_x} \sum_{b=1}^B (x_b - z A v_b)^T (x_b - z A v_b) \right] + \log p(z) \quad (95)$$

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

$$M_k = -\frac{1}{2L} \left[ \sum_{m=1}^{LN} g_k^m \left[ B (C_v^m)^{-1} - (C_v^m)^{-1} \left( \sum_{b=1}^B v^{m,b} v^{(m,b)T} \right) (C_v^m)^{-1} \right] \right] \quad (96)$$