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.

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

$$C_v = \sum_{k=1}^{K} g_k C_k \tag{2}$$

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 α_g and θ_g controlling the sparsity of the prior.

$$p(g) = \mathrm{Gam}(g; \alpha_g, \theta_g) \tag{3}$$

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

$$p(z) = \mathrm{Gam}(z; \alpha_z, \theta_z) \tag{4}$$

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.

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

2 Likelihoods

2.1 Likelihood of g and z

by intuition:

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

by algebraic derivation:

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

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

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

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

2.2 Log-likelihood of the parameters

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

$$p(X \mid C_{1..K}) = \prod_{n=1}^{N} \iint_{-\infty}^{\infty} p(x_n \mid z, g)p(g)p(z)\mathrm{d}g\mathrm{d}z \quad (12)$$

$$(13)$$

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

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

using 6

$$\log p(X \mid C_{1..K}) \approx -N\log L + \sum_{n=1}^{N}\log\sum_{l=1}^{L}\mathcal{N}(x_n; 0, \sigma_x I + z^2 A\left(\sum_{k=1}^{K} g_k U_k^T U_k\right)A^T) \quad (15)$$

$$C_{int} = \sigma_x I + z^2 A\left(\sum_{k=1}^{K} g_k U_k^T U_k\right)A^T \quad (16)$$

$$\mathcal{L}_n^{int} = \sum_{l=1}^{L}\mathcal{N}(x_n; 0, C_{int}) \quad (17)$$

$$\log p(X \mid C_{1..K}) \approx -N\log L + \sum_{n=1}^{N}\log\mathcal{L}_n^{int} \quad (18)$$

using 10

$$p(X \mid C_{1..K}) \approx \left(L\sqrt{\det(A^TA)}\right)^{-N} \prod_{n=1}^{N} \sum_{l=1}^{L} h(z^l)\mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l)) \tag{19}$$

$$\log p(X \mid C_{1..K}) \approx -N(\log L + \frac{1}{2}\log(\det(A^TA))) + \sum_{n=1}^{N} \log\left[\sum_{l=1}^{L} h(z^l)\mathcal{N}(f(x_n, z^l); 0, C(z^l, g^l))\right] \tag{20}$$

$$h^l \equiv h(z^l),\;\; f_n^l \equiv f(x_n, z^l),\;\; C^l \equiv C(z^l, g^l) \tag{21}$$

$$\mathcal{L}_n^l \equiv h^l\mathcal{N}(f_n^l; 0, C^l),\;\; \mathcal{L}_n \equiv \sum_{l=1}^{L} \mathcal{L}_n^l \tag{22}$$

$$\log p(X \mid C_{1..K}) \approx -N(\log L + \frac{1}{2}\log(\det(A^TA))) + \sum_{n=1}^{N} \log \mathcal{L}_n \tag{23}$$

2.2.1 Derivative w.r.t. $U_{1...K}$ based on 10

$$\begin{aligned}\frac{\partial \log p(X \mid C_{1..K})}{\partial\left[U_k\right]_{i,j}} \approx \sum_{n=1}^{N} \frac{\partial \log \mathcal{L}_n}{\partial\left[U_k\right]_{i,j}} = \sum_{n=1}^{N} \frac{1}{\mathcal{L}_n}\frac{\partial \mathcal{L}_n}{\partial\left[U_k\right]_{i,j}} = \sum_{n=1}^{N} \frac{1}{\mathcal{L}_n}\sum_{l=1}^{L}\frac{\partial \mathcal{L}_n^l}{\partial\left[U_k\right]_{i,j}} = \\ = \sum_{n=1}^{N} \frac{1}{\mathcal{L}_n}\sum_{l=1}^{L} \mathrm{Tr}\left[\frac{\partial \mathcal{L}_n^l}{\partial C^l}\frac{\partial C^l}{\partial\left[U_k\right]_{i,j}}\right]\end{aligned} \tag{24}$$

the derivatives in this formula are the following

$$\begin{aligned}\frac{\partial \mathcal{L}_n^l}{\partial C^l} = h^l\frac{\partial}{\partial C^l}\mathcal{N}(f_n^l; 0, C^l) = h^l\mathcal{N}(f_n^l; 0, C^l)\frac{\partial}{\partial C^l}\log\mathcal{N}(f_n^l; 0, C^l) = \\ = -\frac{h^l}{2}\mathcal{N}(f_n^l; 0, C^l)\left[(C^l)^{-1} - (C^l)^{-1}f_n^l(f_n^l)^T(C^l)^{-1}\right]\end{aligned} \tag{25}$$

$$\frac{\partial C(z,g)}{\partial\left[U_k\right]_{i,j}} = g_k\frac{\partial\left(U_k^TU_k\right)}{\partial\left[U_k\right]_{i,j}} = g_k\left(U_k^TJ^{ij} + J^{ji}U_k\right) \equiv g_k\hat{U}_k^{ij} \tag{26}$$

substituting back to the derivative

$$\begin{aligned}\frac{\partial \log p(X \mid C_{1..K})}{\partial\left[U_k\right]_{i,j}} \approx \\ \approx -\frac{1}{2}\sum_{n=1}^{N}\frac{1}{\mathcal{L}_n}\sum_{l=1}^{L} h^lg_k^l\mathcal{N}(f_n^l; 0, C^l)\mathrm{Tr}\left[\left[(C^l)^{-1} - (C^l)^{-1}f_n^l(f_n^l)^T(C^l)^{-1}\right]\hat{U}_k^{ij}\right] = \\ = -\frac{1}{2}\mathrm{Tr}\left[\sum_{n=1}^{N}\left(\frac{1}{\mathcal{L}_n}\sum_{l=1}^{L} h^lg_k^l\mathcal{N}(f_n^l; 0, C^l)\left[(C^l)^{-1} - (C^l)^{-1}f_n^l(f_n^l)^T(C^l)^{-1}\right]\right)\hat{U}_k^{ij}\right]\end{aligned} \tag{27}$$

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

$$M_k = -\frac{1}{2}\sum_{n=1}^{N}\left(\frac{1}{\mathcal{L}_n}\sum_{l=1}^{L} h^l g_k^l \mathcal{N}(f_n^l;0,C^l)\left[(C^l)^{-1}-(C^l)^{-1}f_n^l(f_n^l)^T(C^l)^{-1}\right]\right) \quad (28)$$

$$\frac{\partial \log p(X \mid C_{1..K})}{\partial\left[U_k\right]_{i,j}} \approx \mathrm{Tr}\left[M_k\hat{U}_k^{ij}\right] = \sum_{a=1}^{Dv}\left[M_k\right]_{j,a}\left[U_k\right]_{i,a} + \left[M_k\right]_{a,j}\left[U_k\right]_{i,a} \quad (29)$$

2.2.2 Derivative w.r.t. $U_{1...K}$ based on 6

$$\frac{\partial \log p(X \mid C_{1..K})}{\partial\left[U_k\right]_{i,j}} \approx \sum_{n=1}^{N}\frac{\partial \log \mathcal{L}_n^{int}}{\partial\left[U_k\right]_{i,j}} = \sum_{n=1}^{N}\frac{1}{\mathcal{L}_n^{int}}\frac{\partial \mathcal{L}_n^{int}}{\partial\left[U_k\right]_{i,j}} = \sum_{n=1}^{N}\frac{1}{\mathcal{L}_n^{int}}\sum_{l=1}^{L}\frac{\partial \mathcal{N}(x_n;0,C_{int})}{\partial\left[U_k\right]_{i,j}} =$$
$$= \sum_{n=1}^{N}\frac{1}{\mathcal{L}_n}\sum_{l=1}^{L}\mathrm{Tr}\left[\frac{\partial \mathcal{N}(x_n;0,C_{int})}{\partial C_{int}}\frac{\partial C_{int}}{\partial\left[U_k\right]_{i,j}}\right] \quad (30)$$

$$\frac{\partial \mathcal{N}(x_n;0,C_{int})}{\partial C_{int}} = \frac{\partial}{\partial C_{int}}\mathcal{N}(f_n^l;0,C_{int}) = \mathcal{N}(f_n^l;0,C_{int})\frac{\partial}{\partial C_{int}}\log \mathcal{N}(f_n^l;0,C_{int}) =$$
$$= -\frac{1}{2}\mathcal{N}(f_n^l;0,C_{int})\left[(C_{int})^{-1}-(C_{int})^{-1}f_n^l(f_n^l)^T(C_{int})^{-1}\right] \quad (31)$$

$$\frac{\partial C(z,g)}{\partial\left[U_k\right]_{i,j}} = g_k\frac{\partial\left(U_k^T U_k\right)}{\partial\left[U_k\right]_{i,j}} = g_k\left(U_k^T J^{ij} + J^{ji}U_k\right) \equiv g_k\hat{U}_k^{ij} \quad (32)$$

2.3 Complete-data log-likelihood

2.3.1 Expectation w.r.t. the posterior

2.3.2 Derivative w.r.t. $C_{1...K}$

3 Posteriors

3.1 Full posterior

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

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:

$$\log p(v,g,z \mid x) \sim \log p(x \mid v,z) + \log p(v \mid g) + \log p(g) + \log p(z) =$$
$$= \log \mathcal{N}(x;zAv,\sigma_x I) + \log \mathcal{N}(v;0,C_v) + \log \mathrm{Gam}(g;sh_g,sc_g) + \log \mathrm{Gam}(z;sh_z,sc_z) \quad (34)$$

logarithms of the used pdfs look as follows:

$$\log \mathcal{N}(y;\mu,C) = -\frac{1}{2}\left[\log(2\pi) + \log\det(C) + (y-\mu)^T C^{-1}(y-\mu)\right] \quad (35)$$

$$\log \mathrm{Gam}(y;sh,sc) = \log(1) - \log(\Gamma(sh)) - sh\log(sc) + (sh-1)\log(y) - \frac{y}{sc} \quad (36)$$

3.1.1 Derivative w.r.t. v

3.1.2 Derivative w.r.t. g

3.1.3 Derivative w.r.t. z

3.2 Marginal posterior of g

3.2.1 MAP estimate of g

$$g_{MAP} = \arg\max_g p(g \mid x) = \arg\max_g \frac{p(x \mid g)p(g)}{p(x)} = \arg\max_g p(x \mid g)p(g) \quad (37)$$

$$p(x \mid g) = \int_{-\infty}^{\infty} p(x \mid z,g)p(z)\mathrm{d}z \approx \frac{1}{L}\sum_{l=1}^{L} p(x \mid g, z^l) \quad (38)$$

Appendix

3.3 Merging two Gaussian distributions

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

$$p(x \mid v,z)p(v \mid g) = \mathcal{N}(x; zAv, \sigma_x I)\mathcal{N}(v;0,C_v) \quad (39)$$

The Gaussian over x spelled out is

$$\mathcal{N}(x; zAv, \sigma_x I) = \sqrt{\frac{1}{(2\pi)^{Dx}\sigma_x^{Dx}}}e^{-\frac{1}{2\sigma_x}(x-zAv)^T(x-zAv)} \quad (40)$$

$$-\frac{1}{2}(x-zAv)^T(x-zAv) = -\frac{1}{2}(x^Tx - zv^TA^Tx - zx^TAv + z^2v^TA^TAv) \quad (41)$$

as $v^TA^Tx = (x^TAv)^T$, and both are scalars, thus equal to their transposes, it's also true that $v^TA^Tx = x^TAv$

$$\begin{aligned} -\frac{1}{2}(x-zAv)^T(x-zAv) &= -\frac{1}{2}(x^Tx - 2zx^TAv + z^2v^TA^TAv) = \\ &= -\frac{x^Tx}{2} + zx^TAv - \frac{z^2}{2}v^TA^TAv \end{aligned} \quad (42)$$

we have the identity for any symmetric matrix M and vector b that

$$-\frac{1}{2}v^T M v + b^T v = -\frac{1}{2}(v - M^{-1}b)^T M (v - M^{-1}b) + \frac{1}{2}b^T M^{-1} b \tag{43}$$

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

$$\begin{aligned}
&-\frac{1}{2}(x - zAv)^T (x - zAv) = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+ x)^T z^2 A^T A (v - \frac{1}{z}A^+ x) + \frac{1}{2}(A^T x)^T (A^T A)^{-1} A^T x = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+ x)^T z^2 A^T A (v - \frac{1}{z}A^+ x) + \frac{1}{2}x^T A A^{-1} A^{-T} A^T x = \\
&= -\frac{x^T x}{2} - \frac{1}{2}(v - \frac{1}{z}A^+ x)^T z^2 A^T A (v - \frac{1}{z}A^+ x) + \frac{x^T x}{2} = \\
&= -\frac{1}{2}(v - \frac{1}{z}A^+ x)^T z^2 A^T A (v - \frac{1}{z}A^+ x)
\end{aligned} \tag{44}$$

which implies

$$e^{-\frac{1}{2\sigma_x}(x - zAv)^T (x - zAv)} = e^{-\frac{1}{2}(v - \frac{1}{z}A^+ x^T)^T \frac{z^2}{\sigma_x} A^T A (v - \frac{1}{z}A^+ x)} \tag{45}$$

meaning that

$$\mathcal{N}(x; zAv, \sigma_x I) = \alpha \mathcal{N}(v; \frac{1}{z}A^+ x, \frac{\sigma_x}{z^2}(A^T A)^{-1}) \tag{46}$$

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

$$\sqrt{\frac{1}{(2\pi)^{Dx}\sigma_x^{Dx}}} = \alpha \sqrt{\frac{1}{(2\pi)^{Dv} \det(\frac{\sigma_x}{z^2}(A^T A)^{-1})}} \tag{47}$$

$$\alpha = \sqrt{\frac{(2\pi)^{Dv} \frac{\sigma_x^{D_v}}{z^{2D_v}} \det((A^T A)^{-1})}{(2\pi)^{Dx}\sigma_x^{Dx}}} \tag{48}$$

$$\alpha = \sqrt{\frac{(2\pi)^{Dv}\sigma_x^{D_v}}{(2\pi)^{Dx}\sigma_x^{Dx} z^{2D_v} \det(A^T A)}} \tag{49}$$

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

$$\mathcal{N}(x; zAv, \sigma_x I) = \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \mathcal{N}(v; \frac{1}{z}A^+ x, \frac{\sigma_x}{z^2}(A^T A)^{-1}) \tag{50}$$

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

$$\mathcal{N}(v;\mu_1,C_1)\mathcal{N}(v;\mu_2,C_2)=\mathcal{N}(\mu_1;\mu_2,C_1+C_2)\mathcal{N}(v;\mu_c,C_c) \tag{51}$$

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

$$\begin{aligned}\frac{1}{\sqrt{\det(A^TA)}}\frac{1}{z^{D_v}}\mathcal{N}(v;\frac{1}{z}A^+x,\frac{\sigma_x}{z^2}(A^TA)^{-1})\mathcal{N}(v;0,C_v)=\\ \frac{1}{\sqrt{\det(A^TA)}}\frac{1}{z^{D_v}}\mathcal{N}(\frac{1}{z}A^+x;0,\frac{\sigma_x}{z^2}(A^TA)^{-1}+C_v)\mathcal{N}(v;\mu_c,C_c)\end{aligned} \tag{52}$$

$$C_c=(\frac{z^2}{\sigma_x}(A^TA)+C_v^{-1})^{-1} \tag{53}$$

$$\mu_c=C_c\frac{z}{\sigma_x}(A^TA)A^+x=\frac{z}{\sigma_x}C_cA^Tx \tag{54}$$

3.4 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.

$$\frac{\partial}{\partial y}y^TMy=2My \tag{55}$$

$$\frac{\partial}{\partial y}a^Ty=a \tag{56}$$

$$\frac{\partial}{\partial M}y^TM^{-1}y=-M^{-1}yy^TM^{-1} \tag{57}$$

$$\frac{\partial}{\partial M}\log\det M=M^{-1} \tag{58}$$

$$\frac{\partial}{\partial s}f(M(s))=\mathrm{Tr}\left[\frac{\partial f}{\partial M}\frac{\partial M}{\partial s}\right] \tag{59}$$

$$\frac{\partial}{\partial s}sM=M \tag{60}$$

$$\frac{\partial}{\partial s}f(s)=f(s)\frac{\partial}{\partial s}\log f(s) \tag{61}$$

$$\frac{\partial}{\partial M}\log\mathcal{N}(y;a,M)=M^{-1}-M^{-1}(y-a)(y-a)^TM^{-1} \tag{62}$$

3.5 Equivalence of the likelihood formulas

Expanding 10 yields

$$
\begin{aligned}
p(x \mid z, g) = \frac{1}{\sqrt{\det(A^T A)}} \frac{1}{z^{D_v}} \frac{1}{\sqrt{(2\pi)^{D_v} \det C(z,g)}} e^{-\frac{1}{2}\left(\frac{1}{z}A^+ x\right)^T C^{-1}(z,g)\frac{1}{z}A^+ x} = \\
= \frac{1}{\sqrt{(2\pi z^2)^{D_v} \det(A^T A C(z,g))}} e^{-\frac{1}{2z^2} x^T A^{+T} C^{-1}(z,g) A^+ x}
\end{aligned}
\tag{63}
$$

$$
C(z,g) \equiv \frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v \tag{64}
$$

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

$$
\begin{aligned}
\frac{1}{z^2}\left((A^T A)^{-1} A^T\right)^T \left[\frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v\right]^{-1} (A^T A)^{-1} A^T = \\
= \frac{1}{z^2} A (A^T A)^{-1} \left[\frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v\right]^{-1} (A^T A)^{-1} A^T = \\
= \frac{1}{z^2} A \left[\left[\frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v\right](A^T A)\right]^{-1} (A^T A)^{-1} A^T = \\
= \frac{1}{z^2} A \left[\frac{\sigma_x}{z^2} I + C_v A^T A\right]^{-1} (A^T A)^{-1} A^T = \\
= \frac{1}{z^2} A \left[(A^T A)\left[\frac{\sigma_x}{z^2} I + C_v A^T A\right]\right]^{-1} A^T = \\
= \frac{1}{z^2} A \left[\frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A\right]^{-1} A^T
\end{aligned}
\tag{65}
$$

assuming that A is invertible this is equal to

$$
\begin{aligned}
\frac{1}{z^2}\left[A^{-1}\right]^{-1} \left[\frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A\right]^{-1} \left[A^{-T}\right]^{-1} = \\
\frac{1}{z^2}\left[A^{-1}\right]^{-1} \left[A^{-T}\left[\frac{\sigma_x}{z^2} A^T A + A^T A C_v A^T A\right]\right]^{-1} = \\
\frac{1}{z^2}\left[A^{-1}\right]^{-1} \left[\frac{\sigma_x}{z^2} A + A C_v A^T A\right]^{-1} = \\
\frac{1}{z^2}\left[\left[\frac{\sigma_x}{z^2} A + A C_v A^T A\right] A^{-1}\right]^{-1} = \\
\frac{1}{z^2}\left[\frac{\sigma_x}{z^2} I + A C_v A^T\right]^{-1} = \\
= \left[\sigma_x I + z^2 A C_v A^T\right]^{-1}
\end{aligned}
\tag{66}
$$

under the square root we have

$$
\begin{aligned}
z^{2D_v}\det(A^T A C(z,g)) = z^{2D_v}\det(A^T)\det(A)\det(C(z,g)) = z^{2D_v}\det(A)\det(C(z,g))\det(A^T) \\
= z^{2D_v}\det(A)\det(\frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v)\det(A^T) = z^{2D_v}\det(A)\det(\frac{\sigma_x}{z^2}(A^T A)^{-1} + C_v)\det(A^T) = \\
= z^{2D_v}\det(A)\det(\frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v)\det(A^T) = z^{2D_v}\det\left[A\left[\frac{\sigma_x}{z^2} A^{-1} A^{-T} + C_v\right] A^T\right] = \\
= z^{2D_v}\det\left[\frac{\sigma_x}{z^2} I + A C_v A^T\right] = \det\left[\sigma_x I + z^2 A C_v A^T\right]
\end{aligned}
\tag{67}
$$