Gaussian Scale Mixture Model

Wen-Hao Zhang
Carnegie Mellon University

February 1, 2017

Abstract

This memo presents some calculations of Gaussian mixture model (GSM) and tries to link GSM to explain some functions of neural circucits. In GSM, an observation multivariate Laplacian $\mathbf{x}$ is decomposed into the product of a underlying multivariate Gaussian $\mathbf{y}$ with one or multiple positive random variables z. Two interesting results can be derived in GSM: 1) the elements of $\mathbf{x}$ sharing the same z contribute to the denominator (divisive normalization pool) of the estimate $\mathbf{y}$; 2) the range of correlation in the prior of $\mathbf{y}$ determines the number of elements in the numerator (range of lateral connections) of the estimate $\mathbf{x}$.

1 Generative Model

Suppoe an observation $\mathbf{x}$ is generated by following likelihood function conditioned on underlying variables $\mathbf{y}$ and z,

$$P(\mathbf{x}|\mathbf{y}, z) = \mathcal{N}\left[\mathbf{x}; zA\mathbf{y}, z^2\Sigma_{\mathbf{x}}\right], \tag{1}$$

where $\mathbf{y}$ is an undelrying multivariate Gaussian feature to be inferred, while z is a positive random variable scale the gain of the observation, e.g., luminance level or contrast. For simplicity, I only consider z is a scalar variable, but z can be extended into a vector in general [1, 2].

Unlike previous studies considering a noise-free generative process, i.e., $\Sigma_{\mathbf{x}} = 0$, we consider a multiplicative noise (the variance is proportional to z^2) in this memo for two reasons. First, a non-zero $\Sigma_{\mathbf{x}}$ is essential to derive lateral connections in network implementation for information integration inferring $\mathbf{x}$. Second, the multiplicative noise is supported by the fact that cortical neuron's firing has a multiplicative variability, i.e., Poisson-like variability.

The prior of $\mathbf{y}$ and z are independent with each other. The prior of $\mathbf{y}$ is a multivariate Gaussian distribution,

$$p(\mathbf{y}) = \mathcal{N}[\mathbf{y}; \mu_{\mathbf{y}}, \Sigma_{\mathbf{y}}]. \tag{2}$$

For pursuing an analytical solution, I assume the prior of z taking a Rayleigh distribution,

$$p(z) = z \exp\left(-\frac{z^2}{2}\right). \tag{3}$$

Other forms of priors, e.g., Gamma distribution, were also considered in previous studies ([3, 4]). The usage of Rayleigh distribution in this memo does not change basic results qualitatively, due to similar shape between a Rayleigh and a Gamma distribution.

It can be proved that $p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{y}, z)p(\mathbf{y})p(z)d\mathbf{y}dz$ is a multivariate Laplacian distribution [5, 6].

2 Inference

Let us consider the inference of $\mathbf{y}$ and z given observation $\mathbf{x}$. By using Bayes theorem, we can derive the posterior $p(\mathbf{y}, z|\mathbf{x})$ as

$$p(\mathbf{y}, z|\mathbf{x}) \propto p(\mathbf{y}|\mathbf{x}, z)p(\mathbf{y})p(z). \tag{4}$$

Substituting Eqs. (1-3) into above equation and combing terms about $\mathbf{y}$, we have (see details in Appendix 1)

$$\begin{aligned} p(\mathbf{y}, z|\mathbf{x}) &\propto p(z)\exp\left[-\frac{1}{2}(\mathbf{y}-\mu_{\mathbf{y}})^\top\Sigma_{\mathbf{y}}^{-1}(\mathbf{y}-\mu_{\mathbf{y}}) - \frac{1}{2}(\mathbf{x}-zA\mathbf{y})^\top(z^{-2}\Sigma_{\mathbf{x}}^{-1})(\mathbf{x}-zA\mathbf{y})\right], \\ &\propto p(z)\times\mathcal{N}[\mathbf{x}; (zA)\mu_{\mathbf{y}}, \Sigma_{\mathbf{x}} + (zA)\Sigma_{\mathbf{y}}(zA)^\top]\times\mathcal{N}[\mathbf{y}; \langle\mathbf{y}|\mathbf{x}, z\rangle, \mathrm{Cov}(\mathbf{y}|\mathbf{x}, z)], \end{aligned} \tag{5}$$

where the expression of $\langle\mathbf{y}|\mathbf{x}, z\rangle$ and $\mathrm{Cov}(\mathbf{y}|\mathbf{x}, z)$ can be found in Eq. (15) and Eq. (14) respectively. In the text below, I calculate two marginal posterios $p(z|\mathbf{x})$ and $p(\mathbf{y}|\mathbf{x})$ respectively.

2.1 Inference of $p(z|\mathbf{x})$

$$\begin{aligned} p(z|\mathbf{x}) &= \int p(z, \mathbf{y}|\mathbf{x})d\mathbf{y}, \\ &\propto p(z)\mathcal{N}[\mathbf{x}; (zA)\mu_{\mathbf{y}}, \Sigma_{\mathbf{x}} + (zA)\Sigma_{\mathbf{y}}(zA)^\top] \end{aligned} \tag{6}$$

It seems that $p(z|\mathbf{x})$ in general cannot be analytically solved when $\Sigma_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$ is non-zero. I guess that is the main reason why most of precedent studies have not considered this case. To gain some theoretical insight of a closed form expression of $p(z|\mathbf{x})$, I consider a simple case with $\Sigma_{\mathbf{x}} = 0$ and $\mu_{\mathbf{y}} = 0$ in the below.

In a simplied case that $\Sigma_{\mathbf{x}} = 0$ and $\mu_{\mathbf{y}} = 0$, $p(z|\mathbf{x})$ can be simplified to

$$\begin{aligned} p(z|\mathbf{x})|_{\Sigma_{\mathbf{x}}=0,\mu_{\mathbf{y}}=0} &\propto p(z)\mathcal{N}[\mathbf{x}; 0, (zA)\Sigma_{\mathbf{y}}(zA)^\top], \\ &\propto \frac{1}{z^{n-1}}\exp\left[-\frac{z^2}{2} - \frac{\mathbf{x}^\top(A\Sigma_{\mathbf{y}}A^\top)^{-1}\mathbf{x}}{2z^2}\right], \end{aligned} \tag{7}$$

where n is the dimension of $\mathbf{x}$. Through calculating the normalization factor of $p(z|\mathbf{x})$ (see Appendix 2), $p(z|\mathbf{x})$ is finally solved as

$$p(z|\mathbf{x})|_{\Sigma_{\mathbf{x}}=0,\mu_{\mathbf{y}}=0} = \frac{\lambda^{n/2-1}}{K_{1-n/2}(\lambda)}z^{-(n-1)}\exp\left(-\frac{z^2}{2} - \frac{\lambda}{2z^2}\right), \tag{8}$$

where $\lambda = \mathbf{x}^\top(A\Sigma_{\mathbf{y}}A^\top)^{-1}\mathbf{x}$.

2.2 Inference of $p(\mathbf{y}|\mathbf{x})$

$$p(\mathbf{y}|\mathbf{x}) = \int p(\mathbf{y}|\mathbf{x}, z)p(z|\mathbf{x})dz. \tag{9}$$

Since $p(z|\mathbf{x})$ is already solved in Eq. (8), we next calculate $p(\mathbf{y}|\mathbf{x}, z)$ in the below.

$$\begin{aligned} p(\mathbf{y}|\mathbf{x}, z) &\propto p(\mathbf{x}|\mathbf{y}, z)p(\mathbf{y}|z), \\ &\propto p(\mathbf{x}|\mathbf{y}, z)p(\mathbf{y}), \\ &= \mathcal{N}\left[\mathbf{y}; \langle\mathbf{y}|\mathbf{x}, z\rangle, \mathrm{Cov}(\mathbf{y}|\mathbf{x}, z)\right], \end{aligned} \tag{10}$$

where

$$\mathrm{Cov}(\mathbf{y}|\mathbf{x},z) = \left[(zA)^\top \Sigma_{\mathbf{x}}^{-1}(zA) + \Sigma_{\mathbf{y}}^{-1}\right]^{-1}, \tag{11}$$

$$\langle \mathbf{y}|\mathbf{x},z\rangle = \mathrm{Cov}(\mathbf{y}|\mathbf{x},z)\left[(zA)^\top \Sigma_{\mathbf{x}}^{-1}\mathbf{x} + \Sigma_{\mathbf{y}}^{-1}\mu_{\mathbf{y}}\right]. \tag{12}$$

We see $p(\mathbf{y}|\mathbf{x},z)$ is a multivariate Gaussian distribution, and its mean is a weighted average of $\mathbf{x}$ and $\mu_{\mathbf{y}}$ with the weight proportional to their own reliability (inverse of covariance). This expression is exactly the same as the Bayesian inference widely used in information integration [7, 8]. Moreover, it is worthy to note that off-diagonal elements in $\Sigma_{\mathbf{y}}$ of $p(\mathbf{y})$ denoting the correlation between elements of $\mathbf{y}$, which in turn leads to weighted average between elements of $\mathbf{x}$ in Eq. (12). In network implementation, the weighted average of elements of $\mathbf{x}$ corresponds to lateral connections between neurons.

The closed-form of $p(\mathbf{y}|\mathbf{x})$ seems cannot be obtained in general when $\Sigma_{\mathbf{x}}$ is non-zero. In the special case of $\Sigma_{\mathbf{x}} = 0$, the expression of $p(\mathbf{y}|\mathbf{x},z)$ can be found in Eq. (11) in [1] and Eq. (3.9) in [5].

Appendix

1. Reorganize Gaussian terms

$$
\begin{aligned}
& \exp\left[-\frac{1}{2}(\mathbf{y}-\mu_{\mathbf{y}})^\top\Sigma_{\mathbf{y}}^{-1}(\mathbf{y}-\mu_{\mathbf{y}})-\frac{1}{2}(\mathbf{x}-zA\mathbf{y})^\top(z^{-2}\Sigma_{\mathbf{x}}^{-1})(\mathbf{x}-zA\mathbf{y}\right] \\
= & \exp\left[-\frac{1}{2}(\mathbf{y}-\mu_{\mathbf{y}})^\top\Sigma_{\mathbf{y}}^{-1}(\mathbf{y}-\mu_{\mathbf{y}})-\frac{1}{2}(\mathbf{y}-(zA)^{-1}\mathbf{x})^\top[(zA)^{-1}\Sigma_{\mathbf{x}}(zA)^{-\top}]^{-1}(\mathbf{y}-(zA)^{-1}\mathbf{x})\right].
\end{aligned}
$$

Combing the terms containing $\mathbf{y}$ by using following math tip that

$$
\begin{aligned}
& (\mathbf{x}-\mu_1)^\top\Sigma_1^{-1}(\mathbf{x}-\mu_1)+(\mathbf{x}-\mu_2)^\top\Sigma_2^{-1}(\mathbf{x}-\mu_2) \\
= & (\mathbf{x}-\mu_3)^\top\Sigma_3^{-1}(\mathbf{x}-\mu_3)+(\mu_1-\mu_2)^\top(\Sigma_1+\Sigma_2)^{-1}(\mu_1-\mu_2),
\end{aligned}
$$

where

$$
\begin{aligned}
\Sigma_3 &= (\Sigma_1^{-1}+\Sigma_2^{-1})^{-1}, \\
\mu_3 &= \Sigma_3(\Sigma_1^{-1}\mu_1+\Sigma_2^{-1}\mu_2).
\end{aligned}
$$

We get the final results as

$$
\begin{aligned}
& \exp\left[-\frac{1}{2}(\mathbf{y}-\mu_{\mathbf{y}})^\top\Sigma_{\mathbf{y}}^{-1}(\mathbf{y}-\mu_{\mathbf{y}})-\frac{1}{2}(\mathbf{x}-zA\mathbf{y})^\top(z^{-2}\Sigma_{\mathbf{x}}^{-1})(\mathbf{x}-zA\mathbf{y}\right] \\
\propto & \mathcal{N}[\mathbf{x};(zA)\mu_{\mathbf{y}},\Sigma_{\mathbf{x}}+(zA)\Sigma_{\mathbf{y}}(zA)^\top]\times\mathcal{N}[\mathbf{y};\langle\mathbf{y}|\mathbf{x},z\rangle,\mathrm{Cov}(\mathbf{y}|\mathbf{x},z)], \qquad (13)
\end{aligned}
$$

where

$$
\mathrm{Cov}(\mathbf{y}|\mathbf{x},z) = \left[(zA)^\top\Sigma_{\mathbf{x}}^{-1}(zA)+\Sigma_{\mathbf{y}}^{-1}\right]^{-1}, \qquad (14)
$$

$$
\langle\mathbf{y}|\mathbf{x},z\rangle = \mathrm{Cov}(\mathbf{y}|\mathbf{x},z)\left[(zA)^\top\Sigma_{\mathbf{x}}^{-1}\mathbf{x}+\Sigma_{\mathbf{y}}^{-1}\mu_{\mathbf{y}}\right]. \qquad (15)
$$

2. Normalization factor of $p(z|\mathbf{x})$

Here I calculate the normalization factor, denoted by Z, of $p(z|\mathbf{x})$ in Eq. (7).

$$
Z=\int_0^\infty\frac{1}{z^{n-1}}\exp\left[-\frac{z^2}{2}-\frac{\mathbf{x}^\top(A\Sigma_{\mathbf{y}}A^\top)^{-1}\mathbf{x}}{2z^2}\right]dz \qquad (16)
$$

Denoting $v=z^2$ and $\lambda=\mathbf{x}^\top(A\Sigma_{\mathbf{y}}A^\top)^{-1}\mathbf{x}$, above equation can be transformed to

$$
\begin{aligned}
Z &= \frac{1}{2}\int_0^\infty v^{-n/2}\exp\left(-\frac{v}{2}-\frac{\lambda}{2v}\right)dv, \\
&= \frac{\sqrt{\pi}}{\sqrt{2}\lambda e^\lambda}\int_0^\infty v^{3/2-n/2}\left[\left(\frac{\lambda^2}{2\pi v^3}\right)^{1/2}\exp\left(-\frac{(v-\lambda)^2}{2v}\right)\right]dv. \qquad (17)
\end{aligned}
$$

Recall a inverse Gaussian distribution $\mathrm{IG}[x;\mu,\lambda]$ has probability density function

$$
\mathrm{IG}[x;\mu,\lambda]=\left(\frac{\lambda}{2\pi x^3}\right)^{1/2}\exp\left[-\frac{\lambda(x-\mu)^2}{2\mu^2x}\right], \qquad (18)
$$

and its n-th raw moment is

$$\mathbb{E}_{\mathrm{IG}[x;\mu,\lambda]}[x^n] = e^{\lambda/\mu}\sqrt{\frac{2\lambda}{\pi}}\mu^{n-1/2}K_{1/2-n}\left(\frac{\lambda}{\mu}\right), \tag{19}$$

where $K_n(x)$ is modified Bessel function of the second kind.

Thus, the integral in Eq. (17) can be seen as a $3/2 - n/2$ order raw moment of a inverse Gaussian distribution. By using the result of Eq. (19), we have

$$Z = \lambda^{1-n/2}K_{1-n/2}(\lambda). \tag{20}$$

References

[1] Ruben Coen-Cagli, Peter Dayan, and Odelia Schwartz. Statistical models of linear and nonlinear contextual interactions in early visual processing. In *Advances in neural information processing systems*, pages 369–377, 2009.

[2] Ruben Coen-Cagli, Peter Dayan, and Odelia Schwartz. Cortical surround interactions and perceptual salience via natural scene statistics. *PLoS Comput Biol*, 8(3):e1002405, 2012.

[3] Martin J Wainwright and Eero P Simoncelli. Scale mixtures of gaussians and the statistics of natural images. In *Nips*, pages 855–861, 1999.

[4] Gergő Orbán, Pietro Berkes, József Fiser, and Máté Lengyel. Neural variability and sampling-based probabilistic representations in the visual cortex. *Neuron*, 92(2):530–543, 2016.

[5] Odelia Schwartz, Terrence J Sejnowski, and Peter Dayan. Soft mixer assignment in a hierarchical generative model of natural scene statistics. *Neural Computation*, 18(11):2680–2718, 2006.

[6] Torbjørn Eltoft, Taesu Kim, and Te-Won Lee. On the multivariate laplace distribution. *IEEE Signal Processing Letters*, 13(5):300–303, 2006.

[7] Marc O Ernst and Martin S Banks. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, 415(6870):429–433, 2002.

[8] James J Clark and Alan L Yuille. *Data fusion for sensory information processing systems*, volume 105. Springer Science & Business Media, 2013.