Deriving the Sigmoid Kalman Filter for Neural Mass Models

(Dated: July 20, 2015)

DERIVATION OF COVARIANCE

The derivation of the sigmoid Kalman filter relies on an expansion of the quadratic expression for the covariance, $\hat{P}_t^-$. To make this derivation clearer we have simplified the neural-mass model notation to

$$\xi_t = A\xi_{t-1} + \boldsymbol{\phi}(\xi_{t-1}) + \epsilon_{t-1}, \tag{1}$$

where $\boldsymbol{\phi}(\xi_{t-1}) = B\xi_{t-1} \circ \phi(C\xi_{t-1}, v_0, \varsigma)$. To propagate the covariance, we want to solve for because we use the mean as the estimate $\hat{\xi}_t$

$$\begin{aligned}\hat{P}_t^- =& E\left[\left(\xi_t - \hat{\xi}_t\right)\left(\xi_t - \hat{\xi}_t\right)^\top\right] \\ =& E\left[\left(A\xi_{t-1} + \boldsymbol{\phi}\left(\xi_{t-1}\right) + \epsilon_{t-1} - A\hat{\xi}_{t-1} - E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\right]\right)(\cdot)^\top\right] \\ =& E\left[\left(A\left(\xi_{t-1} - \hat{\xi}_{t-1}\right) + \boldsymbol{\phi}\left(\xi_{t-1}\right) + \epsilon_{t-1} - E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\right]\right)(\cdot)^\top\right],\end{aligned} \tag{2}$$

For notational convenience denote the vectors inside the brackets in order as $a = A\left(\xi_{t-1} - \hat{\xi}_{t-1}\right)$, $b = \boldsymbol{\phi}\left(\xi_{t-1}\right)$, $c = \epsilon_{t-1}$, and $d = E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\right]$ giving

$$\hat{P}_t^- = E\left[(a + b + c - d)(\cdot)^\top\right]. \tag{3}$$

We know

$$E[c] = 0 \tag{4}$$

$$E\left[ca^\top\right] = E\left[cb^\top\right] = E\left[cd^\top\right] = 0 \tag{5}$$

$$E\left[cc^\top\right] = Q \tag{6}$$

$$E\left[aa^\top\right] = A\hat{P}_{t-1}^+ A^\top \tag{7}$$

$$E\left[dd^\top\right] = dd^\top \tag{8}$$

$$E[a] = 0 \tag{9}$$

$$E[b] = d, \tag{10}$$

Using the identities to simplify the expression for $\hat{P}_t$ and substituting back in we get

$$\begin{aligned}\hat{P}_t^- =& A\hat{P}_{t-1}^+ A^\top + Q - E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\right] E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)^\top\right] - A\hat{\xi}_{t-1} E[\boldsymbol{\phi}\left(\xi_{t-1}\right)^\top] - E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\right]\left(A\hat{\xi}_{t-1}\right)^\top \\ &+ E\left[\boldsymbol{\phi}\left(\xi_{t-1}\right)\boldsymbol{\phi}\left(\xi_{t-1}\right)^\top\right] + E\left[(A\xi_{t-1})\boldsymbol{\phi}^\top(\xi_{t-1})\right] + E\left[\boldsymbol{\phi}(\xi_{t-1})\left(A\xi_{t-1}\right)^\top\right].\end{aligned} \tag{11}$$

The results that are derived in this paper are dropping the constant linear terms

$$E\left[\boldsymbol{\phi}\left(\xi\right)\right] = E\left[\xi \circ \phi(\xi, v_0, \varsigma)\right] \tag{12}$$

$$E\left[\boldsymbol{\phi}(\xi)\xi^\top\right] = E\left[\xi \circ \phi(\xi, v_0, \varsigma)\xi^\top\right] \tag{13}$$

$$E\left[\boldsymbol{\phi}\left(\xi\right)\boldsymbol{\phi}\left(\xi\right)^\top\right] = E\left[\phi(\xi, v_0, \varsigma)\phi(\xi, v_0, \varsigma)^\top \circ (\xi\xi^\top)\right] \tag{14}$$

DERIVATION OF SIGMOID TERMS

The following sections will present the results for the terms in Eqns 12 to 14. For ease of notation these derivations are presented for the most simple cases, using the lowest possible order of multivariate Gaussian pdf. The generalization to matrix notation follows from these cases.

A multivariate Gaussian pdf, $p_N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})$ is defined for a random variable $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots & x_N \end{bmatrix}^\top$ with probability density function

$$p_N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}) = f(x_1, \dots, x_N) = \frac{1}{2\pi|\boldsymbol{\sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \tag{15}$$

The mean and covariance are given by

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 & \mu_2 & \dots & \mu_N \end{bmatrix}^\top \tag{16}$$

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \dots & \sigma_{1,N} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{N,1} & \sigma_{N,2} & \dots & \sigma_{N,N} \end{bmatrix} \tag{17}$$

where $\mu_i = \mathbb{E}\left[x_i\right]$, and $\sigma_{i,j} = \sigma_{j,i} = \rho_{ij}\sigma_i\sigma_j$.

The sigmoid function is defined

$$\phi(x) = \frac{1}{\sqrt{2\pi}\varsigma} \int_{-\infty}^{x} \exp\left(-\frac{(z - v_0)^2}{2\varsigma^2}\right) \mathrm{d}z \tag{18}$$

In the following derivation we will use the simple case, where $v_0 = 0$ and $\varsigma = 1$. The extension to the general results for arbitrary v_0 and ς follows trivially from the steps presented here.

Note also, in the following derivations we use placeholder coefficients instead of writing out many terms in full. The full expressions are too long to be written out. At each stage the coefficients were solved and simplified using MATLAB Symbolic Math Toolbox (version 7.10.0, R2013b, The Mathworks Inc., MA, USA).

Derivation of $E\left[\xi \circ \phi(\xi, \upsilon_0, \varsigma)\right]$

We will solve for the simple bivariate case

$$\mathbb{E}\left[f(\mathbf{x})\right] = \mathbb{E}\left[x_1 \phi(x_2)\right] \tag{19}$$

with $\mathbf{x} \sim p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})$ (bivariate normal distribution)

$$\begin{aligned} \mathbb{E}\left[f(\mathbf{x})\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 \phi(x_2) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}x_1\, \mathrm{d}x_2 \qquad (20) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_2} x_1 \exp\left(-\frac{z^2}{2}\right) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}z\, \mathrm{d}x_1\, \mathrm{d}x_2 \end{aligned}$$

We can change the order of integration and shift the z integral (using a change of variable to $\tilde{z}$) giving

$$\mathbb{E}\left[f(\mathbf{x})\right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 \exp\left(-\frac{(\tilde{z} - x_2)^2}{2}\right) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}x_2\, \mathrm{d}x_1\, \mathrm{d}\tilde{z} \tag{21}$$

We expand the quadratic inside the exponent of $p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})$ then collect the x_2 terms inside a single exponent. This gives an integral containing the term of form $\exp(ax_2^2 + bx_2 + c)$, which has the solution

$$\int_{-\infty}^{\infty} \exp\left(ax_2^2 + bx_2 + c\right) \mathrm{d}x_2 = \sqrt{\frac{\pi}{-a}} \exp\left(c - \frac{b^2}{4a}\right), \tag{22}$$

for $\mathrm{Re}(a) < 0$.

We can solve for the terms a, b and c and substitute back into Eqn 31, then perform a similar process to collect the x_1 terms into the canonical form, $x_1 \exp(ax_1^2 + bx_1 + c)$. The solution to this second integral is given by the derivative of Eqn 22 with respect to b, which is;

$$\int_{\infty}^{\infty} x_1 \exp(ax_1^2 + bx_1 + c)\, \mathrm{d}x_1 = b \exp\left(c - \frac{b^2}{4a}\right) \frac{\sqrt{\pi}}{2(-a)^{\frac{3}{2}}}, \tag{23}$$

for $\mathrm{Re}(a) < 0$.

Once we solve for these new a, b and c terms and substitute them back into the remaining integral, we can collect the final expression as factors of $\tilde{z}$ into the form

$$\int_{-\infty}^{0} (d + e\tilde{z}) \exp\left(a\tilde{z}^2 + b\tilde{z} + c\right) \mathrm{d}\tilde{z} \tag{24}$$

This integral can be solved using a combination of two results. The first is

$$\begin{aligned} &d \int_{-\infty}^{0} \exp(a\tilde{z}^2 + b\tilde{z} + c)\,\mathrm{d}\tilde{z} \\ &= d \exp\left(c - \frac{b^2}{4a}\right) \sqrt{\frac{\pi}{-4a}} \left(1 + \operatorname{erf}\left(-\frac{b}{\sqrt{-4a}}\right)\right) \end{aligned} \tag{25}$$

For $\mathrm{Re}(a) < 0$.

The second solution is obtained by taking the derivative of Eqn 25 with respect to b and is given by

$$\begin{aligned} &e \int_{-\infty}^{0} \tilde{z} \exp(a\tilde{z}^2 + b\tilde{z} + c)\,\mathrm{d}\tilde{z} \\ &= e \exp\left(c - \frac{b^2}{4a}\right) \frac{\sqrt{\pi}\sqrt{-a}b \operatorname{erfc}\left(\frac{b}{\sqrt{-4a}}\right) + 2a \exp\left(\frac{b^2}{4a}\right)}{4a^2} \end{aligned} \tag{26}$$

For $\mathrm{Re}(a) < 0$.

The coefficients a, b, c, d and e are;

$$\begin{array}{ll} a = \frac{-1}{2\sigma_{22}+2} & b = \frac{\mu_2}{\sigma_{22}+1} \\ c = -\frac{\mu_1^2}{2\sigma_{22}+2} & d = \frac{\mu_y\sigma_{22}-\mu_2\sigma_{12}+\mu_y}{\sigma_{11}(1+\sigma_{22}-\rho^2\sigma_{22})} \\ e = \frac{\rho\sigma_2}{\sigma_1(1+\sigma_{22}-\rho^2\sigma_{22})} & \end{array} \tag{27}$$

(Note that $a < 0$ and $c - \frac{b^2}{4a} = 0$).

We can combine the solutions in Eqns. 25 and 26 and simplify using the coefficients from Eqns. 27 to give

$$\frac{\mu_1}{2}\operatorname{erf}\left(\frac{\mu_2}{\sqrt{2(\sigma_{22}+1)}}\right) + \frac{\mu_1}{2} + \frac{\sigma_{12}}{\sqrt{2\pi(\sigma_{22}+1)}} \exp\left(-\frac{\mu_2^2}{2(\sigma_{22}+1)}\right) \tag{28}$$

Including v_0 and ς (which does not change the steps followed in the derivation) gives the general form of the solution,

$$\mathbb{E}\left[f(\mathbf{x})\right] = \frac{\mu_1}{2}\operatorname{erf}\left(\frac{\mu_2 - v_0}{\sqrt{2(\sigma_{22}+\varsigma^2)}}\right) + \frac{\mu_1}{2} + \frac{\sigma_{12}}{\sqrt{2\pi(\sigma_{22}+\varsigma^2)}} \exp\left(-\frac{(\mu_2 - v_0)^2}{2(\sigma_{22}+\varsigma^2)}\right) \tag{29}$$

Derivation of $E\left[\xi \circ \phi(\xi, v_0, \varsigma)\xi^\top\right]$

By examining the matrix form of the model we know the highest order case we will need to solve for is

$$\mathbb{E}\left[f(\mathbf{x})\right] = \mathbb{E}\left[x_1 x_2 \phi(x_3)\right] \tag{30}$$

with $\mathbf{x} \sim p_3(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})$

$$\begin{aligned}
\mathbb{E}\left[f(\mathbf{x})\right] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 \phi(x_3) p_3(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{x_3} x_1 x_2 \exp\left(-\frac{z^2}{2}\right) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}z\, \mathrm{d}x_1\, \mathrm{d}x_2\, \mathrm{d}x_3 \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 \exp\left(-\frac{(\tilde{z}-x_3)^2}{2}\right) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma})\, \mathrm{d}x_3\, \mathrm{d}x_2\, \mathrm{d}x_1\, \mathrm{d}\tilde{z}
\end{aligned} \tag{31}$$

The first integral over x_3 takes the canonical form $\int \exp\left(a x_3^2 + b x_3 + c\right) \mathrm{d}x_3$, which can be solved using the result in Eqn. 22. The second integral over x_2 also has a previous canonical form $\int x_2 \exp\left(a x_2^2 + b x_2 + c\right) \mathrm{d}x_2$, which is solved as in Eqn. 23. The third integral over x_1 takes the form $\int (d x_1^2 + e x_1) \exp\left(a x_1^2 + b x_1 + c\right) \mathrm{d}x_1$. This integral can be solved by combining the result from Eqn. 23 with the solution given by:

$$\int_{-\infty}^{\infty} x_1^2 \exp(a x_1^2 + b x_1 + c)\, \mathrm{d}x_1 = \frac{\sqrt{\pi}}{2(-a)^{\frac{3}{2}}}\left(1 - \frac{b^2}{2a}\right) \exp\left(c - \frac{b^2}{4a}\right) \tag{32}$$

Note that Eqn. 32 is obtained from Eqn. 23 by taking the derivative with respect to b. Combining Eqns. 23 and 32 gives

$$\begin{aligned}
&\int_{-\infty}^{\infty} (d x_1^2 + e x_1) \exp(a x_1^2 + b x_1 + c)\, \mathrm{d}x_1 = \\
&\frac{\sqrt{\pi}}{2(-a)^{\frac{3}{2}}}\left(d - \frac{d b^2}{2a} + e b\right) \exp\left(c - \frac{b^2}{4a}\right)
\end{aligned} \tag{33}$$

When the coefficients a-e are solved for and substituted back in the final integral (in $\tilde{z}$) has the form

$$\int_{-\infty}^{0} (d\tilde{z}^2 + e\tilde{z} + f) \exp(a\tilde{z}^2 + b\tilde{z} + c)\, \mathrm{d}\tilde{z} \tag{34}$$

To solve this integral we need the result from Eqns 25 and 26 as well as the second order derivative of Eqn. 25 (or the derivative of Eqn. 26), which is

$$\begin{aligned}
&\int_{-\infty}^{0} \tilde{z}^2 \exp(a\tilde{z}^2 + b\tilde{z} + c)\,\mathrm{d}\tilde{z} \qquad (35)\\
&= \frac{(b^2-2a)\sqrt{\pi}}{8(-a)^{\frac{5}{2}}}\left(1+\operatorname{erf}\left(-\frac{b}{\sqrt{-4a}}\right)\right)\exp\left(c-\frac{b^2}{4a}\right) - \frac{b}{4a^2}\exp(c)
\end{aligned}$$

Combining these we are left with

$$\begin{aligned}
&\int_{-\infty}^{0} (ds^2+es+f)\exp(as^2+bs+c)\,\mathrm{d}s \qquad (36)\\
&= \frac{(b^2d-2ad-2abe+4a^2f)\sqrt{\pi}}{8(-a)^{\frac{5}{2}}}\exp\left(c-\frac{b^2}{4a}\right)\left(1+\operatorname{erf}\left(-\frac{b}{\sqrt{-4a}}\right)\right) + \frac{2ae-db}{4a^2}\exp(c)
\end{aligned}$$

And after solving for the coefficients a-f and simplifying we get to the final result

$$\begin{aligned}
\mathbb{E}\left[f(\mathbf{x})\right] = &\frac{\mu_1\mu_2+\sigma_{12}}{2}\left(1+\operatorname{erf}\left(\frac{\mu_3-v_0}{\sqrt{2(\varsigma^2+\sigma_{33})}}\right)\right) \qquad (37)\\
&-\left(\frac{\sigma_{13}\sigma_{23}(\mu_x-v_0)}{\sqrt{2\pi}(\varsigma^2+\sigma_{33})^{\frac{3}{2}}} - \frac{\sigma_{23}\mu_1+\sigma_{13}\mu_2}{\sqrt{2\pi(\varsigma^2+\sigma_{33})}}\right)\exp\left(-\frac{(\mu_3-v_0)^2}{2(\varsigma^2+\sigma_{33})}\right)
\end{aligned}$$

Derivation of $E\left[\phi(\xi, v_0, \varsigma)\phi(\xi, v_0, \varsigma)^\top \circ (\xi\xi^\top)\right]$

In the filter, the term $E\left[\phi(\xi, v_0, \varsigma)\phi(\xi, v_0, \varsigma)^\top \circ (\xi\xi^\top)\right]$ is approximated as $E\left[\phi(\xi, v_0, \varsigma)\phi(\xi, v_0, \varsigma)^\top\right] \circ (\xi\xi^\top)$ The result for the term $E\left[\phi(\xi, v_0, \varsigma)\phi(\xi, v_0, \varsigma)^\top\right]$ can be calculated by recognizing that the sigmoid function, ϕ can be expressed as the Gaussian cumulative distribution function (cdf), with mean v_0 and variance ς. So the integral can be expressed as

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \Phi\left(\frac{x_1-v_0}{\varsigma}\right)\Phi\left(\frac{x_2-v_0}{\varsigma}\right) p_2(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\sigma})\,\mathrm{d}x_1\,\mathrm{d}x_2 \qquad (38)$$

where $\mathbf{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix}$. We will define two new, independent random variables, z_1 and z_2 that are both normally distributed with mean, v_0 and variance, ς i.e. $z_i \sim p(z|v_0, \varsigma)$. Since z_1 and z_2 are also independent of x_1 and x_2 we can write

$$P(z_1 < x_1, z_2 < x_2 | x_1, x_2) = \Phi\left(\frac{x_1-v_0}{\varsigma}\right)\Phi\left(\frac{x_2-v_0}{\varsigma}\right) \qquad (39)$$

and by law of total probability the unconditional distribution is written

$$\begin{aligned}
P(z_1 < x_1, z_2 < x_2) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(z_1 < x_1, z_2 < x_2 | x_1, x_2) p_2(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\sigma}) \, \mathrm{d}x_1 \, \mathrm{d}x_2 \qquad (40) \\
&= P(z_1 - x_1 < 0, z_2 - x_2 < 0) \\
&= \Phi\left(\frac{\mathbf{z} - \hat{\boldsymbol{\mu}}}{\hat{\boldsymbol{\sigma}}}\right)
\end{aligned}$$

i.e a bivariate Gaussian cdf, with $\mathbf{z} = \begin{bmatrix} z_1 & z_2 \end{bmatrix}$ mean $\hat{\boldsymbol{\mu}} = \begin{bmatrix} \hat{\mu}_1 & \hat{\mu}_2 \end{bmatrix}$ and variance $\hat{\boldsymbol{\sigma}} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} \end{bmatrix}$. These terms can be calculated by recognizing that (since all variables are independent of one another, except for x_1 and x_2)

$$\hat{\mu}_1 = \mathbb{E}[z_1 - x_1] = v_0 - \mu_1 \qquad (41)$$

$$\hat{\mu}_2 = \mathbb{E}[z_2 - x_2] = v_0 - \mu_2 \qquad (42)$$

$$\hat{\sigma}_{11} = \mathrm{var}(z_1 - x_1) = \varsigma^2 + \sigma_{11} \qquad (43)$$

$$\hat{\sigma}_{22} = \mathrm{var}(z_2 - x_2) = \varsigma^2 + \sigma_{22} \qquad (44)$$

$$\begin{aligned}
\hat{\sigma}_{12} &= \mathrm{cov}(z_1 - x_1, z_2 - x_2) \qquad (45) \\
&= \mathrm{cov}(z_1, z_2) + \mathrm{cov}(z_1, x_2) + \mathrm{cov}(x_1, z_2) + \mathrm{cov}(x_1, x_2) \\
&= \mathrm{cov}(x_1, x_2) = \sigma_{12}
\end{aligned}$$