

---

# Supplementary material: Probabilistic amplitude and frequency demodulation

---

**Richard E. Turner\***

Computational and Biological Learning Lab  
Department of Engineering, University of Cambridge  
Trumpington Street, Cambridge, CB2 1PZ, UK  
ret26@cam.ac.uk

**Maneesh Sahani**

Gatsby Computational Neuroscience Unit, University College London  
Alexandra House, 17 Queen Square, London, WC1N 3AR, UK  
maneesh@gatsby.ucl.ac.uk

## 1 Introduction

In this supplementary material we flesh out the application of expectation propagation (EP) to probabilistic amplitude and frequency demodulation (PAFD).

## 2 Forward model

The PAFD generative model draws amplitude variables from an Gaussian auto-regressive process which is constrained to be positive, and phases from an auto-regressive von Mises process (here of second order),

$$p(\mathbf{a}_{1:T}|\lambda_{1:\tau}, \sigma^2) \propto \prod_t \text{Norm}\left(\mathbf{a}_t; \sum_{t'=1}^{\tau} \lambda_{t'} \mathbf{a}_{t-t'}, \sigma^2\right) \mathbf{1}(\mathbf{a}_t \geq 0), \quad (1)$$

$$p(\theta_{1:T}|k_1, k_2) \propto \prod_{t=1}^T \exp(k_1 \cos(\theta_t - \theta_{t-1}) + k_2 \cos(\theta_t - \theta_{t-2})), \quad (2)$$

$$p(y_t|\mathbf{a}_t, \theta_t) = \text{Norm}(y_t; \mathbf{a}_t \cos(\theta_t + \bar{\omega}t), \sigma_y^2). \quad (3)$$

The joint distribution over the amplitudes is a multivariate truncated Gaussian. Therefore, when the von Mises process becomes an independent, uniform distribution ( $k_1 = 0$  and  $k_2 = 0$ ), the model reduces to that of Sell and Slaney with a particular spectral weighting function determined by the dynamical parameter,  $\lambda_{1:\tau}$ .

## 3 Reformulating the model as a constrained Gaussian AR(1) process

The model can be massaged into a form which simplifies the application of EP. First, the distribution over the amplitudes is written as a multi-dimensional Gaussian AR(1) process with an associated positivity constraint. Defining  $\mathbf{a}_t^\top = [\mathbf{a}_{1,t}, \mathbf{a}_{2,t}, \dots, \mathbf{a}_{\tau,t}] = [\mathbf{a}_{1,t}, \mathbf{a}_{1,t-1}, \dots, \mathbf{a}_{1,t-\tau+1}]$ , we have

$$p(\mathbf{a}_{1:T}|\Lambda_{\mathbf{a}}, \Sigma_{\mathbf{a}}) \propto \prod_{t=1}^T \text{Norm}(\mathbf{a}_t; \Lambda_{\mathbf{a}} \mathbf{a}_{t-1}, \Sigma_{\mathbf{a}}) \mathbf{1}(\mathbf{a}_{1,t} \geq 0). \quad (4)$$

---

\*Laboratory for Computational Vision, New York University, 4 Washington Place, New York, NY 10003-6603, USA

The new dynamical parameters are,

$$\Lambda_a = \begin{bmatrix} \lambda_1 & \lambda_2 & \dots & \lambda_{\tau-1} & \lambda_\tau \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}, \quad \Sigma_a = \begin{bmatrix} \sigma_a^2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}. \quad (5)$$

Second, the auto-regressive von Mises process over phases is rewritten as a bivariate Gaussian AR(2) process that is constrained to the unit circle (i.e.  $x_{1,t} = \cos(\theta_t)$  and  $x_{2,t} = \sin(\theta_t)$ ),

$$p(x_{1:2,1:T}) \propto \prod_{t=1}^T \mathbf{1}(x_{1,t}^2 + x_{2,t}^2 = 1) \prod_{m=1}^2 \text{Norm}(x_{m,t}; \lambda_1 x_{m,t-1} + \lambda_2 x_{m,t-2}, \sigma^2), \quad (6)$$

where the parameters of the AR(2) process are related to the parameters of the von Mises distribution by  $\sigma^2 = 1$ ,  $\lambda_1 = \frac{k_1}{1-k_2}$  and  $\lambda_2 = k_2$ .

Alternatively, this bivariate Gaussian AR(2) process can be written as an AR(1) process with four hidden states,  $\mathbf{x}_t^\top = [x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}] = [\cos(\theta_t), \sin(\theta_t), x_{1,t-1}, x_{2,t-1}]$ , that is

$$p(\mathbf{x}_{1:T} | \Lambda_x, \Sigma_x) \propto \prod_{t=1}^T \text{Norm}(\mathbf{x}_t; \Lambda_x \mathbf{x}_{t-1}, \Sigma_x) \mathbf{1}(x_{1,t}^2 + x_{2,t}^2 = 1). \quad (7)$$

The new dynamical parameters relate to the old ones,

$$\Lambda_x = \begin{bmatrix} \lambda_1 & 0 & \lambda_2 & 0 \\ 0 & \lambda_1 & 0 & \lambda_2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad \Sigma_x = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (8)$$

The likelihood must be rewritten in terms of these new variables and it becomes,

$$p(y_t | a_t, \mathbf{x}_t, \mathbf{w}_t, \sigma_y^2) = \text{Norm}(y_t; a_{1,t} \mathbf{w}_t^\top \mathbf{x}_t, \sigma_y^2). \quad (9)$$

Where the time varying weights of the emission distribution are,

$$\mathbf{w}_t^\top = [\cos(\bar{\omega}t) \quad -\sin(\bar{\omega}t) \quad 0 \quad 0]. \quad (10)$$

Finally, to further ease subsequent notation, we can concatenate the amplitude and phase variables  $\mathbf{z}_t^\top = [a_t^\top, \mathbf{x}_t^\top]$  and rewrite the model in terms of this single set of variables,

$$p(\mathbf{z}_{1:T} | \Lambda_z, \Sigma_z) \propto \prod_{t=1}^T \text{Norm}(\mathbf{z}_t; \Lambda_z \mathbf{z}_{t-1}, \Sigma_z) \mathbf{1}(z_{\tau+1,t}^2 + z_{\tau+2,t}^2 = 1) \mathbf{1}(z_{1,t} \geq 0), \quad (11)$$

$$p(y_t | \mathbf{z}_t, \mathbf{w}_t, \sigma_y^2) = \text{Norm}(y_t; z_{1,t} (\mathbf{w}_{1,t} z_{\tau+1,t} + \mathbf{w}_{2,t} z_{\tau+2,t}), \sigma_y^2) \quad (12)$$

The new dynamical parameters are,

$$\Lambda_z = \begin{bmatrix} \Lambda_a & 0 \\ 0 & \Lambda_x \end{bmatrix}, \quad \Sigma_z = \begin{bmatrix} \Sigma_a & 0 \\ 0 & \Sigma_x \end{bmatrix}. \quad (13)$$

This completes the reformulation of the model, which now has constrained Gaussian AR(1) dynamics and a non-linear likelihood which is conditionally Gaussian in  $z_{1,t} = a_{1,t}$  or  $z_{\tau+1:\tau+2} = [x_{1,t}, x_{2,t}]^\top$ , but which is not jointly Gaussian.

In the remainder of this note we use a mixture of the three different notations we have for the model,  $\{a_t, \theta_t\}$ ,  $\{a_t, \mathbf{x}_t\}$ , and  $\mathbf{z}_t$ .

## 4 Factor graph representation of the model

The model can be written in terms of a product of clique potentials in a number of different ways,

$$p(y_{1:T}, \mathbf{z}_{1:T}) \propto \prod_{t=1}^T \text{Norm}(\mathbf{z}_t; \Lambda_z \mathbf{z}_{t-1}, \Sigma_z) p(y_t | \mathbf{z}_t) \mathbf{1}(a_{1,t} \geq 0) \mathbf{1}(x_{1,t}^2 + x_{2,t}^2 = 1), \quad (14)$$

$$= \prod_{t=1}^T \pi_t(\mathbf{z}_t, \mathbf{z}_{t-1}) \psi(a_{1,t}, x_{1,t}, x_{2,t}), \quad (15)$$

$$= \prod_{t=1}^T \Phi_t(\mathbf{z}_t, \mathbf{z}_{t-1}). \quad (16)$$

## 5 Application of Expectation Propagation

Each of the factor graphs described in the previous section suggests a different way to apply EP to the model. Considering numerical stability, implementational convenience, and convergence speed, we choose the representation which breaks apart the potential  $\Phi_t(\mathbf{z}_t, \mathbf{z}_{t-1})$  into the dynamical component and the likelihood/constant component  $\Phi_t(\mathbf{z}_t, \mathbf{z}_{t-1}) = \pi_t(\mathbf{z}_t, \mathbf{z}_{t-1})\psi_t(a_{1,t}, x_{1,t}, x_{2,t})$ . EP then approximates the posterior distribution using a product of forward, backward and constrained-likelihood messages,

$$q(\mathbf{z}_{1:T}) = \prod_{t=1}^T \alpha_t(\mathbf{z}_t) \beta_t(\mathbf{z}_t) \tilde{\psi}_t(a_{1,t}, x_{1,t}, x_{2,t}) = \prod_{t=1}^T q_t(\mathbf{z}_t). \quad (17)$$

All of the messages will be un-normalised Gaussians. The belief will also be un-normalised because the model is only defined up to an unknown normalising constant, which EP will approximate.

The message passing scheme should be interpreted as follows,

1.  $\alpha_t(\mathbf{z}_t)$  is the effect of  $\pi_t(\mathbf{z}_{t-1:t})$  on the belief  $q(\mathbf{z}_t)$
2.  $\beta_t(\mathbf{z}_t)$  is the effect of  $\pi_{t+1}(\mathbf{z}_{t:t+1})$  on the belief  $q(\mathbf{z}_t)$
3.  $\tilde{\psi}_t(a_{1,t}, x_{1,t}, x_{2,t})$  is the effect of the likelihood and the constraints on the belief  $q(\mathbf{z}_t)$

The updates for the messages can be found by removing the messages from  $q(\mathbf{z}_{1:T})$  that correspond to the effect of a particular potential. These messages are replaced by the corresponding potential. The deleted messages are then updated by moment matching the two distributions. We now derive the EP updates using this procedure.

### 5.1 Update for $\alpha_t(\mathbf{z}_t)$ and $\beta_{t-1}(\mathbf{z}_{t-1})$

Removing  $\alpha_t(\mathbf{z}_t)$  and  $\beta_{t-1}(\mathbf{z}_{t-1})$  and replacing them with  $\pi_t(\mathbf{z}_t, \mathbf{z}_{t-1})$  we have,

$$\hat{p}_\pi(\mathbf{z}_{1:T}) = \left( \prod_{t'=1}^{t-2} q_{t'}(\mathbf{z}_{t'}) \right) \alpha_{t-1}(\mathbf{z}_{t-1}) \tilde{\psi}_{t-1}(\mathbf{z}_{t-1}) \pi_t(\mathbf{z}_t, \mathbf{z}_{t-1}) \beta_t(\mathbf{z}_t) \tilde{\psi}_t(\mathbf{z}_t) \left( \prod_{t'=t+1}^T q_{t'}(\mathbf{z}_{t'}) \right), \quad (18)$$

$$= \left( \prod_{t'=1}^{t-2} q_{t'}(\mathbf{z}_{t'}) \right) \hat{p}_\pi(\mathbf{z}_{t-1}, \mathbf{z}_t) \left( \prod_{t'=t+1}^T q_{t'}(\mathbf{z}_{t'}) \right). \quad (19)$$

The moments of belief must be matched to this distribution,  $q(\mathbf{z}_{1:T}) \stackrel{\text{MOM}}{=} \hat{p}_\pi(\mathbf{z}_{1:T})$  by altering the parameters of the messages  $\alpha_t(\mathbf{z}_t)$  and  $\beta_{t-1}(\mathbf{z}_{t-1})$ . From the above, this simplifies to matching the moments of the approximate two-slice conditional,  $q(\mathbf{z}_{t-1}, \mathbf{z}_t) \stackrel{\text{MOM}}{=} \hat{p}_\pi(\mathbf{z}_{t-1}, \mathbf{z}_t)$ . This is simple because all of the distributions are Gaussian.

In more detail, the procedure for updating the forward messages is as follows,

1. Compute the natural parameters of  $\hat{p}_\pi(\mathbf{z}_{t-1}, \mathbf{z}_t)$  by adding the natural parameters of the constituents
2. Convert into moment form (see section 6)
3. Read off the moments for  $\mathbf{z}_t$ . This is equivalent to marginalising  $\mathbf{z}_{t-1}$  to get the one slice marginal  $\hat{p}_\pi(\mathbf{z}_t) = \int d\mathbf{z}_{t-1} \hat{p}_\pi(\mathbf{z}_{t-1}, \mathbf{z}_t)$  and computing the moments of this distribution.
4. Alter the natural parameters of  $\alpha_t(\mathbf{z}_t)$  so that the moments of  $q_t(\mathbf{z}_t)$  match those of  $\hat{p}_\pi(\mathbf{z}_t)$ . This involves converting the moments of  $\hat{p}_\pi(\mathbf{z}_t)$  into natural parameter form, see section 6.

An analogous procedure is followed in the backward message updates.

Both updates require the dynamics potential in natural parameter form,

$$\pi(\mathbf{z}_t, \mathbf{z}_{t-1}) = \text{NormNat} \left( \begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t-1} \end{bmatrix}; \mathbf{c}_t^\pi, \mathbf{c}_{t-1}^\pi, \mathbf{C}_t^\pi \right). \quad (20)$$

where the natural parameters are

$$\mathbf{c}_t^\pi = -\frac{1}{2} \log \det(2\pi\Sigma_z), \quad \mathbf{c}_{t-1}^\pi = \mathbf{0}, \quad \mathbf{C}_t^\pi = -\frac{1}{2} \begin{bmatrix} \Sigma_z^{-1} & -\Sigma_z^{-1}\Lambda_z \\ -\Lambda_z^\top \Sigma_z^{-1} & \Lambda_z^\top \Sigma_z^{-1} \Lambda_z \end{bmatrix}. \quad (21)$$

Using the above the approximate two-slice conditional is,

$$\hat{p}_\pi(\mathbf{z}_t, \mathbf{z}_{t-1}) = \text{NormNat} \left( \begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t-1} \end{bmatrix}; \mathbf{c}_t^{\hat{p}}, \mathbf{c}_{t-1}^{\hat{p}}, \mathbf{C}_t^{\hat{p}} \right). \quad (22)$$

where the natural parameters are

$$\mathbf{c}_t^{\hat{p}} = -\frac{1}{2} \log \det(2\pi \Sigma_z) + \tilde{\mathbf{c}}_t + \tilde{\mathbf{c}}_{t-1}, \quad \mathbf{c}_t^{\hat{p}} = \begin{bmatrix} \tilde{\mathbf{c}}_t \\ \tilde{\mathbf{c}}_{t-1} \end{bmatrix}, \quad \mathbf{C}_t^{\hat{p}} = \begin{bmatrix} -\frac{1}{2} \Sigma_z^{-1} + \tilde{\mathbf{C}}_t & \frac{1}{2} \Sigma_z^{-1} \Lambda_z \\ \frac{1}{2} \Lambda_z^T \Sigma_z^{-1} & -\frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z + \tilde{\mathbf{C}}_{t-1} \end{bmatrix}.$$

where the natural parameters denoted with a tilde are formed from the incoming messages at  $t$  and  $t-1$  by adding their natural parameters,

$$\tilde{\mathbf{c}}_t = \mathbf{c}_t^\beta + \mathbf{c}_t^\psi, \quad \tilde{\mathbf{c}}_{t-1} = \mathbf{c}_{t-1}^\alpha + \mathbf{c}_{t-1}^\psi, \quad \tilde{\mathbf{C}}_t = \mathbf{C}_t^\beta + \mathbf{C}_t^\psi, \quad (23)$$

$$\tilde{\mathbf{c}}_{t-1} = \mathbf{c}_{t-1}^\alpha + \mathbf{c}_{t-1}^\psi, \quad \tilde{\mathbf{C}}_t = \mathbf{C}_t^\beta + \mathbf{C}_t^\psi, \quad \tilde{\mathbf{C}}_{t-1} = \mathbf{C}_{t-1}^\alpha + \mathbf{C}_{t-1}^\psi. \quad (24)$$

First we consider the updates of the forward messages. In order to compute these we integrate  $\mathbf{z}_{t-1}$  ( $\hat{p}_\pi(\mathbf{z}_t) = \int d\mathbf{z}_{t-1} \hat{p}_\pi(\mathbf{z}_t, \mathbf{z}_{t-1})$ ), which by the results in section 7, yields,

$$\hat{p}_\pi(\mathbf{z}_t) = \text{NormNat}(\mathbf{z}_t; \mathbf{c}_t, \mathbf{C}_t), \quad (25)$$

where

$$\mathbf{C}_t = -\frac{1}{2} \Sigma_z^{-1} + \tilde{\mathbf{C}}_t - \frac{1}{4} \Sigma_z^{-1} \Lambda_z (\tilde{\mathbf{C}}_{t-1} - \frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z)^{-1} \Lambda_z^T \Sigma_z^{-1} = \tilde{\mathbf{C}}_t - (2\Sigma_z - \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T)^{-1} \quad (26)$$

Where we have used the matrix inversion lemma,  $(P + BRB^T)^{-1} = P^{-1} - P^{-1}B(R^{-1} + B^T P^{-1}B)^{-1}B^T P^{-1}$ , with  $P = -2\Sigma_z$ ,  $B = \Lambda_z$ ,  $R = \tilde{\mathbf{C}}_{t-1}^{-1}$ .

$$\mathbf{c}_t = \tilde{\mathbf{c}}_t - \frac{1}{2} \Sigma_z^{-1} \Lambda_z (\tilde{\mathbf{C}}_{t-1} - \frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z)^{-1} \tilde{\mathbf{c}}_{t-1} = \tilde{\mathbf{c}}_t - (2\Sigma_z - \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T)^{-1} \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \tilde{\mathbf{c}}_{t-1}. \quad (27)$$

Where we have used  $P^{-1}B(R^{-1} + B^T P^{-1}B)^{-1} = (P + BRB^T)^{-1}BR$  with  $P = 2\Sigma_z$ ,  $B = \Lambda_z$  and  $R = -\tilde{\mathbf{C}}_{t-1}^{-1}$ .

$$\begin{aligned} \mathbf{c}_t &= -\frac{1}{2} \log \det 2\Sigma_z + \tilde{\mathbf{c}}_t - \frac{1}{4} \tilde{\mathbf{c}}_{t-1}^T (\tilde{\mathbf{C}}_{t-1} - \frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z)^{-1} \tilde{\mathbf{c}}_{t-1} - \frac{1}{2} \log \det (\frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z - \tilde{\mathbf{C}}_{t-1}), \\ &= \tilde{\mathbf{c}}_t + \tilde{\mathbf{c}}_{t-1} - \frac{1}{2} \log \det (-\tilde{\mathbf{C}}_{t-1}) - \frac{1}{2} \log \det (2\Sigma_z - \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T) \\ &\quad - \frac{1}{4} \tilde{\mathbf{c}}_{t-1}^T (\tilde{\mathbf{C}}_{t-1}^{-1} + \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T (2\Sigma_z - \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T)^{-1} \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1}) \tilde{\mathbf{c}}_{t-1}. \end{aligned} \quad (28)$$

Where we have used the matrix determinant lemma,  $\det(R + B^T P^{-1}B) = \det(P + BRB^T) \det(P^{-1}) \det(R)$  and the matrix inversion lemma.

We update the forward messages by matching moments  $\hat{p}_\pi(\mathbf{z}_t) \stackrel{\text{MOM}}{=} q(\mathbf{z}_t)$ . Because everything is linear and Gaussian, this can be done exactly by matching the natural parameters.

$$\mathbf{C}_t^\alpha = -\frac{1}{2} \mathbf{V}_t^{-1}, \quad (29)$$

$$\mathbf{c}_t^\alpha = -\frac{1}{2} \mathbf{V}_t^{-1} \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \tilde{\mathbf{c}}_{t-1}, \quad (30)$$

$$\mathbf{c}_t^\alpha = \tilde{\mathbf{c}}_{t-1} - \frac{1}{2} \log \det (-\tilde{\mathbf{C}}_{t-1}) - \frac{1}{2} \log \det 2\mathbf{V}_t - \frac{1}{4} \tilde{\mathbf{c}}_{t-1}^T (\tilde{\mathbf{C}}_{t-1}^{-1} + \frac{1}{2} \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T \mathbf{V}_t^{-1} \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1}) \tilde{\mathbf{c}}_{t-1}. \quad (31)$$

where

$$\mathbf{V}_t = \Sigma_z - \frac{1}{2} \Lambda_z \tilde{\mathbf{C}}_{t-1}^{-1} \Lambda_z^T \quad (32)$$

Next we consider the backward messages. In order to compute these we integrate  $\mathbf{z}_t$  ( $\hat{p}_\pi(\mathbf{z}_{t-1}) = \int d\mathbf{z}_t \hat{p}_\pi(\mathbf{z}_t, \mathbf{z}_{t-1})$ ), which by the results in section 7, yields,

$$\hat{p}_\pi(\mathbf{z}_{t-1}) = \text{NormNat}(\mathbf{z}_{t-1}; \mathbf{c}_{t-1}, \mathbf{C}_{t-1}), \quad (33)$$

where

$$\begin{aligned} \mathbf{C}_{t-1} &= -\frac{1}{2} \Lambda_z^T \Sigma_z^{-1} \Lambda_z + \tilde{\mathbf{C}}_{t-1} - \frac{1}{4} \Lambda_z^T \Sigma_z^{-1} (\tilde{\mathbf{C}}_t - \frac{1}{2} \Sigma_z^{-1})^{-1} \Sigma_z^{-1} \Lambda_z = \tilde{\mathbf{C}}_{t-1} - \Lambda_z^T (2\Sigma_z - \tilde{\mathbf{C}}_t^{-1})^{-1} \Lambda_z, \\ \mathbf{c}_{t-1} &= \tilde{\mathbf{c}}_{t-1} - \frac{1}{2} \Lambda_z^T \Sigma_z^{-1} (\tilde{\mathbf{C}}_t - \frac{1}{2} \Sigma_z^{-1})^{-1} \tilde{\mathbf{c}}_t = \tilde{\mathbf{c}}_{t-1} - \Lambda_z^T (2\Sigma_z - \tilde{\mathbf{C}}_t^{-1})^{-1} \tilde{\mathbf{C}}_t^{-1} \tilde{\mathbf{c}}_t, \end{aligned} \quad (34)$$

$$\begin{aligned} \mathbf{c}_{t-1} &= \tilde{\mathbf{c}}_t + \tilde{\mathbf{c}}_{t-1} - \frac{1}{4} \tilde{\mathbf{c}}_t^T (\tilde{\mathbf{C}}_t - \frac{1}{2} \Sigma_z^{-1})^{-1} \tilde{\mathbf{c}}_t - \frac{1}{2} \log \det (\frac{1}{2} \Sigma_z^{-1} - \tilde{\mathbf{C}}_t) - \frac{1}{2} \log \det (2\Sigma_z), \\ &= \tilde{\mathbf{c}}_t + \tilde{\mathbf{c}}_{t-1} - \frac{1}{4} \tilde{\mathbf{c}}_t^T \left( \tilde{\mathbf{C}}_t^{-1} - \tilde{\mathbf{C}}_t^{-1} (\tilde{\mathbf{C}}_t^{-1} - 2\Sigma_z)^{-1} \tilde{\mathbf{C}}_t^{-1} \right) \tilde{\mathbf{c}}_t - \frac{1}{2} \log \det (2\Sigma_z - \tilde{\mathbf{C}}_t^{-1}) - \frac{1}{2} \log \det (-\tilde{\mathbf{C}}_t) \end{aligned} \quad (35)$$

This leads to the backward recursions,

$$C_{t-1}^\beta = -\frac{1}{2}\Lambda_z^\top U_t^{-1}\Lambda_z, \quad \mathbf{c}_{t-1}^\beta = -\frac{1}{2}\Lambda_z^\top U_t^{-1}\tilde{C}_t^{-1}\tilde{\mathbf{c}}_t, \quad (36)$$

$$\mathbf{c}_{t-1}^\beta = \tilde{\mathbf{c}}_t - \frac{1}{4}\tilde{\mathbf{c}}_t^\top \left( \tilde{C}_t^{-1} + \frac{1}{2}\tilde{C}_t^{-1}U_t^{-1}\tilde{C}_t^{-1} \right) \tilde{\mathbf{c}}_t - \frac{1}{2}\log \det(2U_t) - \frac{1}{2}\log \det(-\tilde{C}_t). \quad (37)$$

where

$$U_t = \Sigma_z - \frac{1}{2}\tilde{C}_t^{-1}. \quad (38)$$

The forward and backward messages reduce to the Kalman Smoother recursions when the likelihood is an unconstrained Gaussian.

## 5.2 Update for $\tilde{\psi}_t(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t})$

In order to update the constrained likelihood messages, we divide out the new message  $\tilde{\psi}_t(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t})$ , and replace it with the true potential yielding the moment matching update,

$$q(\mathbf{x}_t) \stackrel{\text{MOM}}{=} \hat{p}_\psi(\mathbf{z}_t) = \alpha_t(\mathbf{z}_t)\beta_t(\mathbf{z}_t)\psi_t(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t}) \quad (39)$$

The true potential  $\psi_t(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t})$  is a truncated Gaussian in the amplitude variable  $\mathbf{a}_t$ ,

$$\psi_t(\mathbf{a}_t, \theta_t) = \text{NormNat}\left(\mathbf{a}_t; \mathbf{c}_t^\psi, \mathbf{c}_t^\psi, C_t^\psi\right) \mathbf{1}(\mathbf{a}_t \geq 0). \quad (40)$$

where,

$$\mathbf{c}_t^\psi = -\frac{1}{2}\log 2\pi\sigma_y^2 - \frac{1}{2\sigma_y^2}\mathbf{y}_t^2, \quad \mathbf{c}_t^\psi = \frac{1}{\sigma_y^2}\mathbf{y}_t \cos(\bar{\omega}t + \theta_t), \quad C_t^\psi = -\frac{1}{2\sigma_y^2} \cos^2(\bar{\omega}t + \theta_t). \quad (41)$$

This observation motivates the following scheme to compute the moments of  $\hat{p}_\psi(\mathbf{z}_t)$ . First we can marginalise all of the variables except the amplitude ( $\mathbf{a}_t$ ) and the phasor variables ( $\mathbf{x}_{1,t}$  and  $\mathbf{x}_{2,t}$ , or equivalently  $\theta$ ). This is simple because they are Gaussian. We can then compute the necessary moments by first integrating over the amplitude variable (the integrals involve error-function) and then numerically integrating over the phase variable (e.g. by gridding up the space).

In more detail first we compute the natural parameters of the product of the incoming messages,  $\alpha_t(\mathbf{z}_t)\beta_t(\mathbf{z}_t) = \text{NormNat}(\mathbf{z}_t; \mathbf{c}_t^{\alpha\beta}, \mathbf{c}_t^{\alpha\beta}, C_t^{\alpha\beta})$ ,

$$\mathbf{c}_t^{\alpha\beta} = \mathbf{c}_t^\alpha + \mathbf{c}_t^\beta, \quad \mathbf{c}_t^{\alpha\beta} = \mathbf{c}_t^\alpha + \mathbf{c}_t^\beta, \quad C_t^{\alpha\beta} = C_t^\alpha + C_t^\beta. \quad (42)$$

Next we marginalise over all variables except  $\mathbf{a}_t, \mathbf{x}_{1,t}$  and  $\mathbf{x}_{2,t}$  using the result in section 7 and write the result as a function of the amplitude variable,

$$\hat{p}_\psi(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t}) = \psi_t(\mathbf{a}_t, \theta_t) \int d\{\mathbf{z}_t \neq \mathbf{a}_t, \theta_t\} \text{NormNat}(\mathbf{z}_t; \mathbf{c}_t^{\alpha\beta}, \mathbf{c}_t^{\alpha\beta}, C_t^{\alpha\beta}), \quad (43)$$

$$= \text{NormNat}(\mathbf{a}_t; \mathbf{c}^*(\theta), \mathbf{c}^*(\theta), C^*(\theta)) \mathbf{1}(\mathbf{a}_t \geq 0). \quad (44)$$

Next we compute the zeroth, first and second moments of the amplitude variable  $\mathbf{a}_t$  using section 8.

$$\text{mom}_0(\theta_t) = \int d\mathbf{a}_t \hat{p}_\psi(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t}), \quad (45)$$

$$\text{mom}_1(\theta_t) = \int d\mathbf{a}_t \mathbf{a}_t \hat{p}_\psi(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t}), \quad (46)$$

$$\text{mom}_2(\theta_t) = \int d\mathbf{a}_t \mathbf{a}_t^2 \hat{p}_\psi(\mathbf{a}_t, \mathbf{x}_{1,t}, \mathbf{x}_{2,t}). \quad (47)$$

Finally we compute the circular moments by integrating over the phase variable using, for example, using the rectangle rule,

$$z = \int d\theta_t \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \text{mom}_0(n\Delta_\theta), \quad (48)$$

$$\langle a_t \rangle = \int d\theta_t \text{mom}_1(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \text{mom}_1(n\Delta_\theta), \quad (49)$$

$$\langle \sin(\theta_t) \rangle = \int d\theta_t \sin(\theta_t) \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \sin(n\Delta_\theta) \text{mom}_0(n\Delta_\theta), \quad (50)$$

$$\langle \cos(\theta_t) \rangle = \int d\theta_t \cos(\theta_t) \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \cos(n\Delta_\theta) \text{mom}_0(n\Delta_\theta), \quad (51)$$

$$\langle a_t^2 \rangle = \int d\theta_t \text{mom}_2(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \text{mom}_2(n\Delta_\theta), \quad (52)$$

$$\langle a_t \cos(\theta) \rangle = \int d\theta_t \cos(\theta_t) \text{mom}_1(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \cos(n\Delta_\theta) \text{mom}_1(n\Delta_\theta), \quad (53)$$

$$\langle a_t \sin(\theta) \rangle = \int d\theta_t \sin(\theta_t) \text{mom}_1(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \sin(n\Delta_\theta) \text{mom}_1(n\Delta_\theta), \quad (54)$$

$$\langle \cos^2(\theta) \rangle = \int d\theta_t \cos^2(\theta_t) \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \cos^2(n\Delta_\theta) \text{mom}_0(n\Delta_\theta), \quad (55)$$

$$\langle \sin^2(\theta) \rangle = \int d\theta_t \sin^2(\theta_t) \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \sin^2(n\Delta_\theta) \text{mom}_0(n\Delta_\theta), \quad (56)$$

$$\langle \sin(\theta) \cos(\theta) \rangle = \int d\theta_t \sin(\theta_t) \cos(\theta_t) \text{mom}_0(\theta_t) \approx \frac{\Delta_\theta}{N} \sum_{n=1}^N \sin(n\Delta_\theta) \cos(n\Delta_\theta) \text{mom}_0(n\Delta_\theta). \quad (57)$$

Having established these moments, we update the constrained likelihood potential natural parameters by converting the moments to natural parameters and subtracting off the natural parameters of the incoming messages,  $\alpha$  and  $\beta$ .

## 6 Natural and moment parameterisations of a Gaussian

The natural parameter form of the Gaussian is,

$$\text{NormNat}(\mathbf{x}; \mathbf{c}, \mathbf{C}) = \exp(\mathbf{c}^\top \mathbf{x} + \mathbf{x}^\top \mathbf{C} \mathbf{x}) \quad (58)$$

The natural parameter form is useful because multiplication of two Gaussians involves adding the natural parameters,

$$\text{NormNat}(\mathbf{x}; \mathbf{c}_1 + \mathbf{c}_2, \mathbf{C}_1 + \mathbf{C}_2) = \text{NormNat}(\mathbf{x}; \mathbf{c}_1, \mathbf{C}_1) \text{NormNat}(\mathbf{x}; \mathbf{c}_2, \mathbf{C}_2). \quad (59)$$

The ‘moment’ form of the Gaussian is,

$$\text{NormMom}(\mathbf{x}; z, \mu, \Sigma) = \frac{z}{\sqrt{(\det(2\pi\Sigma))}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu)\right) \quad (60)$$

The moment form is useful because the moments are given by,

$$z = \int d\mathbf{x} \text{NormMom}(\mathbf{x}; z, \mu, \Sigma), \quad (61)$$

$$z\mu = \int d\mathbf{x} \mathbf{x} \text{NormMom}(\mathbf{x}; z, \mu, \Sigma), \quad (62)$$

$$z(\Sigma + \mu\mu^\top) = \int d\mathbf{x} \mathbf{x}\mathbf{x}^\top \text{NormMom}(\mathbf{x}; z, \mu, \Sigma). \quad (63)$$

The relationship between the natural and moment parameterisations is,

$$\Sigma = -\frac{1}{2}\mathbf{C}^{-1}, \quad \mu = -\frac{1}{2}\mathbf{C}^{-1}\mathbf{c}, \quad z = \frac{(\pi)^{D/2}}{\sqrt{\det(-\mathbf{C})}} \exp\left(\mathbf{c} - \frac{1}{4}\mathbf{c}^\top \mathbf{C}^{-1}\mathbf{c}\right) \quad (64)$$

And the opposite mapping is,

$$C = -\frac{1}{2}\Sigma^{-1}, \quad \mathbf{c} = \Sigma^{-1}\mu, \quad c = \log z - \frac{1}{2}\log \det(2\pi\Sigma) - \frac{1}{2}\mu^\top \Sigma^{-1}\mu. \quad (65)$$

## 7 Marginalising a subset of Gaussian variables

Computation of the moments takes the form of an integral of an un-normalised Gaussian with a non-linear potential. Whilst the Gaussian component depends on all the variables, denoted here  $\{\mathbf{x}, \mathbf{y}\}$ , the non-linear potential often depends on just a subset,  $\mathbf{x}$ . This section shows to integrate out the Gaussian subset,  $\mathbf{y}$ .

The moments we need to compute are,

$$\langle 1 \rangle = \int d\mathbf{x} d\mathbf{y} \text{NormNat} \left( \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; c_{\mathbf{xy}}, \mathbf{c}_{\mathbf{xy}}, C_{\mathbf{xy}} \right) \gamma(\mathbf{x}) = \int d\mathbf{x} m(\mathbf{x}) \gamma(\mathbf{x}), \quad (66)$$

$$\langle \mathbf{x} \rangle = \int d\mathbf{x} d\mathbf{y} \mathbf{x} \text{NormNat} \left( \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; c_{\mathbf{xy}}, \mathbf{c}_{\mathbf{xy}}, C_{\mathbf{xy}} \right) \gamma(\mathbf{x}) = \int d\mathbf{x} \mathbf{x} m(\mathbf{x}) \gamma(\mathbf{x}), \quad (67)$$

$$\langle \mathbf{x}\mathbf{x}^\top \rangle = \int d\mathbf{x} d\mathbf{y} \mathbf{x}\mathbf{x}^\top \text{NormNat} \left( \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; c_{\mathbf{xy}}, \mathbf{c}_{\mathbf{xy}}, C_{\mathbf{xy}} \right) \gamma(\mathbf{x}) = \int d\mathbf{x} \mathbf{x}\mathbf{x}^\top m(\mathbf{x}) \gamma(\mathbf{x}). \quad (68)$$

These integrals have been broken down into an initial integral over  $\mathbf{y}$  and a subsequent integral over  $\mathbf{x}$ . The initial integral is,

$$m(\mathbf{x}) = \int d\mathbf{y} \text{NormNat} \left( \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; c_{\mathbf{xy}}, \mathbf{c}_{\mathbf{xy}}, C_{\mathbf{xy}} \right). \quad (69)$$

In order to compute this integral we rewrite the natural parameterised Gaussian over  $\mathbf{x}$  and  $\mathbf{y}$  as a moment parameterised Gaussian over  $\mathbf{y}$ . Using the block-partitioned form of the parameters,

$$\mathbf{c}_{\mathbf{xy}} = \begin{bmatrix} \mathbf{c}_x \\ \mathbf{c}_y \end{bmatrix}, \quad C_{\mathbf{xy}} = \begin{bmatrix} C_{\mathbf{xx}} & C_{\mathbf{xy}} \\ C_{\mathbf{xy}}^\top & C_{\mathbf{yy}} \end{bmatrix}. \quad (70)$$

Using these partitioned forms,

$$\text{NormNat} \left( \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}; c_{\mathbf{xy}}, \mathbf{c}_{\mathbf{xy}}, C_{\mathbf{xy}} \right) = \text{NormNat}(\mathbf{x}; \mathbf{c}^*, \mathbf{c}^*, C^*) \text{NormMom}(\mathbf{y}; 1, \mu_{\mathbf{y}|\mathbf{x}}, \Sigma_{\mathbf{y}|\mathbf{x}}) \quad (71)$$

The moment-parameters are given by,

$$\mu_{\mathbf{y}|\mathbf{x}} = -C_{\mathbf{yy}}^{-1}(\mathbf{c}_y/2 + C_{\mathbf{xy}}^\top \mathbf{x}), \quad \Sigma_{\mathbf{y}|\mathbf{x}} = -\frac{1}{2}C_{\mathbf{yy}}^{-1} \quad (72)$$

and the natural parameters are given by

$$\mathbf{c}^* = \mathbf{c}_{\mathbf{xy}} - \frac{1}{4}\mathbf{c}_y^\top C_{\mathbf{yy}}^{-1} \mathbf{c}_y + \frac{D}{2} \log(\pi) - \frac{1}{2} \log \det(-C_{\mathbf{yy}}), \quad (73)$$

$$C^* = C_{\mathbf{xx}} - C_{\mathbf{xy}} C_{\mathbf{yy}}^{-1} C_{\mathbf{xy}}^\top. \quad (74)$$

Where  $D$  is the number of variables which have been marginalised. This new form makes the integral simple to compute,

$$m(\mathbf{x}) = \text{NormNat}(\mathbf{x}; \mathbf{c}^*, \mathbf{c}^*, C^*). \quad (75)$$

## 8 Moments of a truncated Gaussian

The constrained likelihood message update requires the moments of a truncated Gaussian,

$$\text{mom}_k = \int_0^\infty da a^k \text{NormMom}(a; z_a, \mu_a, \sigma_a^2) \quad \forall k = \{0, 1, 2\}. \quad (76)$$

Taking these integrals in turn,

$$\text{mom}_0 = z_a \int_0^\infty da \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{1}{2\sigma_a^2}(a - \mu_a)^2\right) \quad (77)$$

$$= \frac{z_a}{\sqrt{\pi}} \int_{-\frac{\mu_a}{\sqrt{2}\sigma_a}}^\infty du \exp(-u^2), \quad (78)$$

$$= \frac{z_a}{\sqrt{\pi}} \left( \int_0^{\frac{\mu_a}{\sqrt{2}\sigma_a}} du \exp(-u^2) + \int_0^\infty du \exp(-u^2) \right), \quad (79)$$

$$= \frac{z_a}{2} \left( \text{erf}\left(\frac{\mu_a}{\sqrt{2}\sigma_a}\right) + 1 \right), \quad (80)$$

Where we have converted the integral into a standard form using  $u = \frac{1}{\sqrt{2}\sigma_a} (a - \mu_a)$  and we have made use of the symmetry of the integral, and define the error-function as  $\text{erf}(x) = \frac{2}{\pi} \int_0^x du \exp(-u^2)$ .

$$\text{mom}_1 = z_a \int_0^\infty da \, a \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{1}{2\sigma_a^2} (a - \mu_a)^2\right), \quad (81)$$

$$= z_a \int_0^\infty da \, (a - \mu_a) \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{1}{2\sigma_a^2} (a - \mu_a)^2\right) + \mu_a \text{mom}_0, \quad (82)$$

$$= \frac{z_a \sigma_a}{\sqrt{2\pi}} \int_{\frac{\mu_a}{2\sigma_a^2}}^\infty da \exp(-u) + \mu_a \text{mom}_0, \quad (83)$$

$$= -\frac{z_a \sigma_a}{\sqrt{2\pi}} [\exp(-u)]_{\frac{\mu_a}{2\sigma_a^2}}^\infty + \mu_a \text{mom}_0, \quad (84)$$

$$= \frac{z_a \sigma_a}{\sqrt{2\pi}} \exp\left(-\frac{\mu_a^2}{2\sigma_a^2}\right) + \mu_a \text{mom}_0. \quad (85)$$

where we made the substitution,  $u = \frac{1}{2\sigma_a^2} (a - \mu_a)^2$ .

$$\text{mom}_2 = z_a \int_0^\infty da \, a^2 \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{1}{2\sigma_a^2} (a - \mu_a)^2\right), \quad (86)$$

$$= z_a \int_0^\infty da \, (a - \mu_a)^2 \frac{1}{\sqrt{2\pi\sigma_a^2}} \exp\left(-\frac{1}{2\sigma_a^2} (a - \mu_a)^2\right) + 2\mu_a \text{mom}_1 - \mu_a^2 \text{mom}_0, \quad (87)$$

$$= \frac{2\sigma_a^2 z_a}{\sqrt{\pi}} \int_{-\frac{\mu_a}{\sqrt{2}\sigma_a}}^\infty du \, u^2 \exp(-u^2) + 2\mu_a \text{mom}_1 - \mu_a^2 \text{mom}_0, \quad (88)$$

$$= \frac{\sigma_a^2 z_a}{\sqrt{\pi}} \left( -[u \exp(-u^2)]_{-\frac{\mu_a}{\sqrt{2}\sigma_a}}^\infty + \int_{-\frac{\mu_a}{\sqrt{2}\sigma_a}}^\infty du \exp(-u^2) \right) + 2\mu_a \text{mom}_1 - \mu_a^2 \text{mom}_0, \quad (89)$$

$$= -\frac{\sigma_a z_a \mu_a}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_a^2} \mu_a^2\right) + \sigma_a^2 \text{mom}_0 + 2\mu_a \text{mom}_1 - \mu_a^2 \text{mom}_0, \quad (90)$$

$$= \mu_a \text{mom}_1 + \sigma_a^2 \text{mom}_0. \quad (91)$$

Here we have integrated by parts using  $\frac{dv}{du} = u \exp(-u^2)$  and  $v = -\frac{1}{2} \exp(-u^2)$ .