Information form filtering and smoothing for Gaussian linear dynamical systems

Scott W. Linderman   Matthew J. Johnson

November 13, 2016

The *information form* of the Gaussian density of $x \in \mathbb{R}^n$ is defined as,

$$p(x \mid J, h) = \exp\left\{-\frac{1}{2}x^{\mathsf{T}} J x + h^{\mathsf{T}} x - \log Z\right\}, \tag{1}$$

where

$$\log Z = \frac{1}{2} h^{\mathsf{T}} J^{-1} h - \frac{1}{2}\log|J| + \frac{n}{2}\log 2\pi. \tag{2}$$

The standard formulation is recovered by the transformations, $\Sigma = J^{-1}$, and $\mu = J^{-1}h$. The advantage of working in the information form is that it corresponds to the natural parameterization of the Gaussian distribution, and mean field variational inference is considerably easier with this form.

In order to perform Kalman filtering and smoothing, we must be able to perform two operations: *conditioning* and *marginalization*.

Conditioning If,

$$p(x) = \mathcal{N}(x \mid J, h) \tag{3}$$
$$p(y \mid x) \propto \mathcal{N}(x \mid J_{\mathsf{obs}}, h_{\mathsf{obs}}) \tag{4}$$

then,

$$p(x \mid y) = \mathcal{N}(x \mid J + J_{\mathsf{obs}}, h + h_{\mathsf{obs}}). \tag{5}$$

Marginalization If,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} J_{11} & J_{12} \\ J_{12}^{\mathsf{T}} & J_{22} \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}\right), \tag{6}$$

then,

$$x_2 \sim \mathcal{N}(J_{22} - J_{21}J_{11}^{-1}J_{21}^{\mathsf{T}},\ h_2 - J_{21}J_{11}^{-1}h_1) \tag{7}$$

Proof. We use the following integral identity,

$$\int \exp\left\{-\frac{1}{2}x^\mathsf{T} J x + h^\mathsf{T} x\right\} \mathrm{d}x = \exp\left\{\frac{1}{2}h^\mathsf{T} J^{-1} h - \frac{1}{2}\log|J| + \frac{n}{2}\log 2\pi\right\}. \tag{8}$$

For $x = [x_1, x_2]^\mathsf{T}$, we marginalize out x_1 via,

$$\int \exp\left\{-\frac{1}{2}\begin{bmatrix} x_1^\mathsf{T} & x_2^\mathsf{T}\end{bmatrix}\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22}\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix} + \begin{bmatrix} h_1^\mathsf{T} & h_2^\mathsf{T}\end{bmatrix}\begin{bmatrix} x_1 \\ x_2\end{bmatrix} - \log Z\right\} \mathrm{d}x_1. \tag{9}$$

Rearranging terms, we have,

$$\begin{aligned}
&\exp\left\{-\frac{1}{2}x_2^\mathsf{T} J_{22} x_2 + h_2^\mathsf{T} x_2 - \log Z\right\} \int \exp\left\{-\frac{1}{2}x_1^\mathsf{T} J_{11} x_1 + (h_1 - J_{12}x_2)^\mathsf{T} x_1\right\} \mathrm{d}x_1 && (10)\\
&= \exp\left\{-\frac{1}{2}x_2^\mathsf{T} J_{22} x_2 + h_2^\mathsf{T} x_2 - \log Z\right\} \exp\left\{\frac{1}{2}(h_1 - J_{12}x_2)^\mathsf{T} J_{11}^{-1}(h_1 - J_{12}x_2) - \frac{1}{2}\log|J_{11}| + \frac{n}{2}\log 2\pi\right\} && (11)\\
&= \exp\left\{-\frac{1}{2}x_2^\mathsf{T}(J_{22} - J_{21}J_{11}^{-1}J_{12})x_2 + (h_2 - J_{21}J_{11}^{-1}h_1)^\mathsf{T} x_2 - \log Z\right\} \\
&\qquad \times \exp\left\{\frac{1}{2}h_1^\mathsf{T} J_{11}^{-1} h_1 - \frac{1}{2}\log|J_{11}| + \frac{n}{2}\log 2\pi\right\}. && (12)
\end{aligned}$$

We recognize this as a Gaussian potential on x_2 of the form,

$$\begin{aligned}
p(x_2) &= \exp\left\{-\frac{1}{2}x_2^\mathsf{T}\widetilde{J}_2 x_2 + \widetilde{h}_2^\mathsf{T} x_2 - \log\widetilde{Z}_2\right\} && (13)\\
\widetilde{J}_2 &= J_{22} - J_{21}J_{11}^{-1}J_{12} && (14)\\
\widetilde{h}_2 &= h_2 - J_{21}J_{11}^{-1}h_1 && (15)\\
\log\widetilde{Z}_2 &= \log Z - \frac{1}{2}h_1^\mathsf{T} J_{11}^{-1} h_1 + \frac{1}{2}\log|J_{11}| - \frac{n}{2}\log 2\pi. && (16)
\end{aligned}$$

□

Filtering, Sampling, and Smoothing

By interleaving these two steps we can filter, sample, and smooth the latent states in a linear dynamical system. Take the model,

$$\begin{aligned}
x_1 &\sim \mathcal{N}(\mu_1, Q_1) && (17)\\
x_{t+1} &\sim \mathcal{N}(A_t x_t + B_t u_t, Q_t) && (18)\\
y_t &\sim \mathcal{N}(C_t x_t + D_t u_t, R_t). && (19)
\end{aligned}$$

In information form, the initial distribution is,

$$x_1 \sim \mathcal{N}(J = Q_1^{-1}, h = Q_1^{-1}\mu_1). \tag{20}$$

The dynamics are given by,

$$p(x_{t+1} \mid x_t) \propto \mathcal{N}\left(\begin{bmatrix} x_t \\ x_{t+1} \end{bmatrix} \,\middle|\, \begin{bmatrix} J_{11} & J_{12} \\ J_{12}^\mathsf{T} & J_{22} \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right), \tag{21}$$

with,

$$J_{11} = A_t^\mathsf{T} Q_t^{-1} A_t, \quad J_{12} = -A_t^\mathsf{T} Q_t^{-1} \quad J_{22} = Q_t^{-1} \quad h_1 = -u_t^\mathsf{T} B_t^\mathsf{T} Q_t^{-1} A_t \quad h_2 = u_t^\mathsf{T} B_t Q_t^{-1}. \tag{22}$$

Finally, the observations are given by,

$$p(y_t \mid x_t) \propto \mathcal{N}(x_t \mid J_{\mathsf{obs}}, h_{\mathsf{obs}}) \tag{23}$$

with

$$J_{\mathsf{obs}} = C_t^\mathsf{T} R_t^{-1} C_t \quad h_{\mathsf{obs}} = (y_t - D_t u_t)^\mathsf{T} R_t^{-1} C_t \tag{24}$$

Filtering

We seek the conditional distribution, $p(x_t \mid y_{1:t})$, which will be Gaussian. We begin with the initial distribution,

$$p(x_1) = \mathcal{N}(x_1 \mid J_{1|0}, h_{1|0}). \tag{25}$$

Assume, inductively, that $x_t \mid y_{1:t-1} \sim \mathcal{N}(J_{t|t-1}, h_{t|t-1})$. Conditioning on the t-th observation yields, Conditioned on the first observation,

$$\begin{aligned} p(x_t \mid y_{1:t}) &= \mathcal{N}(x_t \mid J_{t|t}, h_{t|t}), && (26) \\ J_{t|t} &= J_{t|t-1} + J_{\mathsf{obs}} && (27) \\ h_{t|t} &= h_{t|t-1} + h_{\mathsf{obs}}. && (28) \end{aligned}$$

Then, we predict the next latent state by writing the joint distribution of x_t and x_{t+1} and marginalizing out x_t.

$$\begin{aligned} p(x_{t+1} \mid y_{1:t}) &= p(x_t \mid y_{1:t})\, p(x_{t+1} \mid x_t) && (29) \\ &= \mathcal{N}(x_t \mid J_{t+1|t}, h_{t+1|t}) && (30) \\ J_{t+1|t} &= J_{22} - J_{21}(J_{t|t} + J_{11})^{-1} J_{21}^\mathsf{T} && (31) \\ h_{t+1|t} &= h_2 - J_{21}(J_{t|t} + J_{11})^{-1}(h_{t|t} + h_1) && (32) \end{aligned}$$

. This completes one iteration and provides the input to the next. To start the recursion, we initialize,

$$J_{1|0} = \Sigma_{\mathsf{init}}^{-1}, \quad h_{1|0} = \Sigma_{\mathsf{init}}^{-1} \mu_{\mathsf{init}}. \tag{33}$$

Marginal Likelihood

This filtering algorithm corresponds to message passing in a chain-structured Gaussian graphical model. To compute the marginal likelihood, $p(y_{1:T})$, we observe that it is the normalization constant of the graphical model,

$$
\begin{aligned}
p(x_{1:T}) &= \frac{1}{Z} \prod_{t=1}^{T} \psi(x_t) \prod_{t=1}^{T-1} \psi(x_t, x_{t+1}), && (34)\\
\psi(x_1) &= p(x_1)\, p(y_1 \mid x_1) && (35)\\
\psi(x_t) &= p(y_t \mid x_t) \qquad \text{for } t = 2 \dots T && (36)\\
\psi(x_t, x_{t+1}) &= p(x_{t+1} \mid x_t) \qquad \text{for } t = 1 \dots T-1. && (37)
\end{aligned}
$$

To compute the normalizaion constant Z, we recursively eliminate nodes via message passing.

Base case. Let us work backward from the final step in which we are left with a graph with a single, unnormalized Gaussian potential.

$$
p(x_T) = \frac{1}{Z}\psi(x_T) = \frac{1}{Z} \exp\left\{-\frac{1}{2}x_T^\mathsf{T} J_T x_T + h_T^\mathsf{T} x_T - \log Z_T\right\} \tag{38}
$$

To compute the normalizing constant, Z, we integrate over x_T and use the normalizing constant for Gaussian distributions:

$$
\begin{aligned}
Z &= \int \exp\left\{-\frac{1}{2}x_T^\mathsf{T} J_T x_T + h_T^\mathsf{T} x_T - \log Z_T\right\} \mathrm{d}x_T && (39)\\
&= \exp\{-\log Z_T\} \int \exp\left\{-\frac{1}{2}x_T^\mathsf{T} J_T x_T + h_T^\mathsf{T} x_T\right\} \mathrm{d}x_T && (40)\\
&= \exp\left\{-\log Z_T + \frac{1}{2}h_T^\mathsf{T} J_T^{-1} h_T - \frac{1}{2}\log|J_T| + \frac{n}{2}\log 2\pi\right\} && (41)
\end{aligned}
$$

Condition Step. In the second to last step, we have two potentials, one from the dynamics induced by the preceding step and one from the observation:

$$
\begin{aligned}
p(x_T) &= \frac{1}{Z}\psi_{\mathsf{dyn}}(x_T)\, \psi_{\mathsf{obs}}(x_T) && (42)\\
&= \exp\left\{-\frac{1}{2}x_T^\mathsf{T} J_{T|T-1} x_T + h_{T|T-1}^\mathsf{T} x_T - \log Z_{T|T-1}\right\} \exp\left\{-\frac{1}{2}x_T^\mathsf{T} J_{\mathsf{obs}} x_T + h_{\mathsf{obs}}^\mathsf{T} x_T - \log Z_{\mathsf{obs}}\right\} && (43)
\end{aligned}
$$

We reduce this to the base case by simply summing the natural parameters and log normalizers,

$$
\begin{aligned}
J_T &= J_{T|T-1} + J_{\mathsf{obs}} && (44)\\
h_T &= h_{T|T-1} + h_{\mathsf{obs}} && (45)\\
\log Z_T &= \log Z_{T|T-1} + \log Z_{\mathsf{obs}}. && (46)
\end{aligned}
$$

Predict Step. Now consider a model with two latent states.

$$
p(x_{T-1}, x_T) = \frac{1}{Z}\psi(x_{T-1})\, \psi_{\mathsf{dyn}}(x_{T-1}, x_T)\, \psi_{\mathsf{obs}}(x_T) \tag{47}
$$

We will eliminate the variable x_{T-1} by marginalizing, or integrating it out.

$$p(x_T) = \frac{1}{Z}\psi_{\mathsf{obs}}(x_T) \int \psi(x_{T-1})\, \psi_{\mathsf{dyn}}(x_{T-1}, x_T)\mathrm{d}x_{T-1} \tag{48}$$

The integrand is of the form,

$$\psi(x_{T-1})\, \psi_{\mathsf{dyn}}(x_{T-1}, x_T) \tag{49}$$

$$= \exp\left\{-\frac{1}{2}x_{T-1}^{\mathsf{T}} J_{T-1} x_{T-1} + h_{T-1}^{\mathsf{T}} x_{T-1} - \log Z_{T-1}\right\} \tag{50}$$

$$\times \exp\left\{-\frac{1}{2}\begin{bmatrix} x_{T-1}^{\mathsf{T}} & x_T^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}\begin{bmatrix} x_{T-1} \\ x_T \end{bmatrix} + \begin{bmatrix} h_1^{\mathsf{T}} & h_2^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} x_{T-1} \\ x_T \end{bmatrix} - \log Z_{\mathsf{dyn}}\right\} \tag{51}$$

$$= \exp\left\{-\frac{1}{2}\begin{bmatrix} x_{T-1}^{\mathsf{T}} & x_T^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} J_{11} + J_{T-1} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}\begin{bmatrix} x_{T-1} \\ x_T \end{bmatrix} + \begin{bmatrix} (h_1 + h_{T-1})^{\mathsf{T}} & h_2^{\mathsf{T}} \end{bmatrix}\begin{bmatrix} x_{T-1} \\ x_T \end{bmatrix} - (\log Z_{T-1} + \log Z_{\mathsf{dyn}})\right\} \tag{52}$$

Appealing to the marginalization propositions above, this implies,

$$\int \psi(x_{T-1})\, \psi_{\mathsf{dyn}}(x_{T-1}, x_T)\, \mathrm{d}x_{T-1} = \exp\left\{-\frac{1}{2}x_T^{\mathsf{T}} J_{T|T-1} x_T + h_{T|T-1}^{\mathsf{T}} x_T - \log Z_{T|T-1}\right\}, \tag{53}$$

where

$$\log Z_{T|T-1} = \\ \log Z_{T-1} + \log Z_{\mathsf{dyn}} - \frac{1}{2}(h_1 + h_{T-1}^{\mathsf{T}})(J_{11} + J_{T-1})^{-1}(h_1 + h_{T-1}^{\mathsf{T}}) + \frac{1}{2}\log|J_{11} + J_{T-1}| - \frac{n}{2}\log 2\pi. \tag{54}$$

Thus, the log normalizer passed into the condition step is an accumulation of log normalizers from previous time steps ($\log Z_{T-1}$), the log normalizer of the dynamics potential ($\log Z_{\mathsf{dyn}}$), and a term that comes from marginalizing out the local variable x_{T-1}. Once we have computed $\log Z_{T|T-1}$, it is passed into the condition step and then into the final integration to compute the marginal likelihood.

This process of predicting and conditioning is recursively applied, marginalizing out variables one at a time, starting with x_1 and ending with x_T. At the end of this procedure, we are left with the marginal likelihood of the observations.

Standard Form Marginal likelihood

The marginal likelihood of the observed data is given by,

$$\log p(y_{1:T} \mid u_{1:T}) = \sum_{t=1}^{T} \log p(y_t \mid y_{1:t-1}, u_{1:t}) \tag{55}$$

$$= \sum_{t=1}^{T} \log \mathcal{N}(y_t \mid C\mu_{t|t-1} + Du_t, S_t), \tag{56}$$

where

$$S_t = C\Sigma_{t|t-1}C^{\mathsf{T}} + R_t \tag{57}$$
$$\mu_{t|t-1} = J_{t|t-1}^{-1}h_{t|t-1} \tag{58}$$
$$\Sigma_{t|t-1} = J_{t|t-1}^{-1} \tag{59}$$

Expanding the above, we have,

$$\log p(y_{1:T} \,|\, u_{1:T}) = \sum_{t=1}^{T}\left[-\frac{N}{2}\log 2\pi - \frac{1}{2}\log\left|S_t\right| - \frac{1}{2}(y_t - C\mu_{t|t-1} - Du_t)^{\mathsf{T}} S_t^{-1}(y_t - C\mu_{t|t-1} - Du_t)\right] \tag{60}$$

Backward Sampling

Having computed $J_{t|t}$ and $h_{t|t}$, we the proceed backward in time to draw a joint sample of the latent states. Given $J_{t|t}$, $h_{t|t}$, and x_{t+1}, we have,

$$p(x_t \,|\, y_{1:t}, x_{t+1}) \propto p(x_t \,|\, y_{1:T})\, p(x_{t+1} \,|\, x_t) \tag{61}$$
$$\propto \mathcal{N}(x_t \,|\, J_{t|t}, h_{t|t})\, \mathcal{N}(x_t \,|\, J_{11},\; h_1 - x_{t+1}^{\mathsf{T}} J_{21}) \tag{62}$$
$$\propto \mathcal{N}(x_t \,|\, J_{t|t} + J_{11},\; h_{t|t} + h_1 - x_{t+1}^{\mathsf{T}} J_{21}) \tag{63}$$

We sample x_t from this conditional, then use it to sample x_{t-1}, and repeat until we reach x_1.

Rauch-Tung-Striebel Smoothing

Next we seek the conditional distribution given all the data, $p(x_t \,|\, y_{1:T})$. This will again be Gaussian, and we will call its parameters $J_{t|T}$ and $h_{t|T}$. Assume, inductively, that we have computed $J_{t+1|T}$ and $h_{t+1|T}$. We show how to compute the parameters for time t.

From the Markov properties of the model and the conditional distribution derived above, we have,

$$p(x_t \,|\, x_{t+1}, y_{1:T}) = \mathcal{N}(x_t \,|\, J_{t|t} + J_{11},\; h_{t|t} + h_1 - J_{12}x_{t+1}). \tag{64}$$

Expanding, taking care to note that x_{t+1} appears in the normalizing constant, yields,

$$\begin{aligned} p(x_t \,|\, x_{t+1}, y_{1:T}) = \exp\Big\{ &-\frac{1}{2}x_t^{\mathsf{T}}(J_{t|t} + J_{11})x_t + (h_{t|t} + h_1)^{\mathsf{T}} x_t - x_{t+1}^{\mathsf{T}} J_{12} x_t \\ &- \frac{1}{2}x_{t+1}^{\mathsf{T}} J_{12}^{\mathsf{T}}(J_{t|t} + J_{11})^{-1}J_{12}x_{t+1} + (h_{t|t} + h_1)^{\mathsf{T}}(J_{t|t} + J_{11})^{-1}J_{12}x_{t+1} \\ &- \frac{1}{2}(h_{t|t} + h_1)^{\mathsf{T}}(J_{t|t} + J_{11})^{-1}(h_{t|t} + h_1)\Big\} \end{aligned} \tag{65}$$

Now consider the joint distribution of x_t and x_{t+1} given all the data,

$$p(x_t, x_{t+1} \,|\, y_{1:T}) = p(x_t \,|\, x_{t+1}, y_{1:T})p(x_{t+1} \,|\, y_{1:T}) \tag{66}$$
$$\propto \mathcal{N}\left(\begin{bmatrix} x_t \\ x_{t+1} \end{bmatrix} \,\middle|\, \begin{bmatrix} \widetilde{J}_{11} & \widetilde{J}_{12} \\ \widetilde{J}_{12}^{\mathsf{T}} & \widetilde{J}_{22} \end{bmatrix}, \begin{bmatrix} \widetilde{h}_1 \\ \widetilde{h}_2 \end{bmatrix}\right), \tag{67}$$

with,

$$
\begin{aligned}
\widetilde{J}_{11} &= J_{t|t} + J_{11} && (68)\\
\widetilde{J}_{12} &= J_{12} && (69)\\
\widetilde{J}_{22} &= J_{t+1|T} + J_{12}^{\mathsf{T}}(J_{t|t} + J_{11})^{-1} J_{12} && (70)\\
\widetilde{h}_{1} &= h_{t|t} + h_{1} && (71)\\
\widetilde{h}_{2} &= h_{t+1|T} + (h_{t|t} + h_{1})^{\mathsf{T}}(J_{t|t} + J_{11})^{-1} J_{12}. && (72)
\end{aligned}
$$

Recall that,

$$
\begin{aligned}
J_{t+1|t} &= J_{22} - J_{21}(J_{t|t} + J_{11})^{-1} J_{21}^{\mathsf{T}} && (73)\\
h_{t+1|t} &= h_{2} - J_{21}(J_{t|t} + J_{11})^{-1}(h_{t|t} + h_{1}). && (74)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\widetilde{J}_{22} &= J_{t+1|T} - J_{t+1|t} + J_{22} && (75)\\
\widetilde{h}_{2} &= h_{t+1|T} - h_{t+1|t} + h_{2}. && (76)
\end{aligned}
$$

Finally, marginalize,

$$
\begin{aligned}
p(x_t \mid y_{1:T}) &= \mathcal{N}(x_t \mid \widetilde{J}_{11} - \widetilde{J}_{12}\widetilde{J}_{22}^{-1}\widetilde{J}_{12}^{\mathsf{T}},\ \widetilde{h}_{1} - \widetilde{J}_{12}\widetilde{J}_{22}^{-1}\widetilde{h}_{2}) && (77)\\
&= \mathcal{N}(x_t \mid J_{t|T}, h_{t|T}). && (78)
\end{aligned}
$$

Substituting the simplified forms above yields,

$$
\begin{aligned}
J_{t|T} &= J_{t|t} + J_{11} - J_{12}(J_{t+1|T} - J_{t+1|t} + J_{22})^{-1} J_{12}^{\mathsf{T}} && (79)\\
h_{t|T} &= h_{t|t} + h_{1} - J_{12}(J_{t+1|T} - J_{t+1|t} + J_{22})^{-1}(h_{t+1|T} - h_{t+1|t} + h_{2}). && (80)
\end{aligned}
$$

Working out marginal likelihood in a simple example

$$
\begin{aligned}
p(x) &= \frac{1}{Z}\mathcal{N}(x \mid 0, 1)\mathcal{N}(1 \mid x, \sigma^2) && (81)\\
&= \frac{1}{Z}\exp\left\{-\frac{1}{2}x^2 - \frac{1}{2}\log 2\pi\right\}\exp\left\{-\frac{1}{2}\frac{(x-1)^2}{\sigma^2} - \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2\right\} && (82)\\
&= \frac{1}{Z}\exp\left\{-\frac{1}{2}(x^2 + \frac{x^2}{\sigma^2}) - \frac{1}{2}\frac{-2x}{\sigma^2} - \frac{1}{2}\frac{1}{\sigma^2} - \log 2\pi - \frac{1}{2}\log\sigma^2\right\} && (83)\\
&= \frac{1}{Z}\exp\left\{-\frac{1}{2}(1 + \frac{1}{\sigma^2})x^2 + \frac{x}{\sigma^2} - \frac{1}{2\sigma^2} - \log 2\pi - \frac{1}{2}\log\sigma^2\right\} && (84)
\end{aligned}
$$

This implies that the normalization constant is,

$$Z = \exp\left\{-\log 2\pi - \frac{1}{2\sigma^2} + \frac{1}{2}(\frac{1}{\sigma^2}(1+\frac{1}{\sigma^2})^{-1}\frac{1}{\sigma^2}) - \frac{1}{2}\log|1+\frac{1}{\sigma^2}| + \frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2\right\} \tag{85}$$

$$= \exp\left\{-\frac{1}{2}\log 2\pi - \frac{1}{2\sigma^2} + \frac{1}{2\sigma^2}\frac{1}{1+\sigma^2} - \frac{1}{2}\log(1+\sigma^2) + \frac{1}{2}\log\sigma^2 - \frac{1}{2}\log\sigma^2\right\} \tag{86}$$

$$= \exp\left\{-\frac{1}{2}\log 2\pi - \frac{1}{2}\frac{1}{1+\sigma^2} - \frac{1}{2}\log(1+\sigma^2)\right\} \tag{87}$$

From the standard marginal likelihood, we know that this is like,

$$\log p(y=1) = \log \mathcal{N}(1\,|\,0, 1+\sigma^2) \tag{88}$$

$$= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log(1+\sigma^2) - \frac{1}{2}\frac{1^2}{1+\sigma^2} \tag{89}$$

$$\tag{90}$$

Thankfully, all checks out!

Working out marginal likelihood in an input example

$$p(x) = \frac{1}{Z}\mathcal{N}(x\,|\,0,1)\,\mathcal{N}(1\,|\,x+d,1) \tag{91}$$

$$= \frac{1}{Z}\exp\left\{-\frac{1}{2}x^2 - \frac{1}{2}\log 2\pi\right\}\exp\left\{-\frac{1}{2}(x+d-1)^2 - \frac{1}{2}\log 2\pi\right\} \tag{92}$$

$$= \frac{1}{Z}\exp\left\{-\frac{1}{2}(x^2+x^2) - \frac{1}{2}2x(d-1) - \frac{1}{2}(d-1)^2 - \log 2\pi\right\} \tag{93}$$

$$= \frac{1}{Z}\exp\left\{-\frac{1}{2}2x^2 + x(1-d) - \frac{1}{2}(d-1)^2 - \log 2\pi\right\} \tag{94}$$

This implies that the normalization constant is,

$$Z = \exp\left\{-\log 2\pi - \frac{1}{2}(d-1)^2 + \frac{1}{2}\frac{(1-d)^2}{2} - \frac{1}{2}\log|2| + \frac{1}{2}\log 2\pi\right\} \tag{95}$$

$$= \exp\left\{-\frac{1}{2}\log 2\pi - \frac{1}{4}(d-1)^2 - \frac{1}{2}\log|2|\right\} \tag{96}$$

From the standard marginal likelihood, we know that this is like,

$$\log p(y=1) = \log \mathcal{N}(1\,|\,d, 2) \tag{97}$$

$$= -\frac{1}{2}\log 2\pi - \frac{1}{2}\log 2 - \frac{1}{2}\frac{(1-d)^2}{2} \tag{98}$$

$$\tag{99}$$

Thankfully, all checks out!