EP: A quick reference

Thomas Minka

February 15, 2008 (originally Oct 2004)

1 Gaussian EP

You want to approximate the term $f_i(\mathbf{w})$ by

$$\tilde{f}_i(\mathbf{w}) = s_i \exp(-\frac{1}{2}(\mathbf{w} - \mathbf{m}_i)^{\mathrm{T}} \mathbf{V}_i^{-1} (\mathbf{w} - \mathbf{m}_i)) \tag{1}$$

To remove a term:

$$\begin{aligned}
q(\mathbf{w}) &= s\, \mathcal{N}(\mathbf{w}; \mathbf{m}, \mathbf{V}) && (2)\\
q^{\setminus i}(\mathbf{w}) &= \mathcal{N}(\mathbf{w}; \mathbf{m}^{\setminus i}, \mathbf{V}^{\setminus i}) \propto \frac{q(\mathbf{w})}{\tilde{f}_i(\mathbf{w})} && (3)\\
\mathbf{V}^{\setminus i} &= (\mathbf{V}^{-1} - \mathbf{V}_i^{-1})^{-1} && (4)\\
\mathbf{m}^{\setminus i} &= \mathbf{V}^{\setminus i}(\mathbf{V}^{-1}\mathbf{m} - \mathbf{V}_i^{-1}\mathbf{m}_i) && (5)
\end{aligned}$$

The ADF equations come from the following relations (obtained from integration-by-parts):

$$\begin{aligned}
Z_i(\mathbf{m}^{\setminus i}, \mathbf{V}^{\setminus i}) &= \int_{\mathbf{w}} f(\mathbf{w}) q^{\setminus i}(\mathbf{w})\; d\mathbf{w} && (6)\\
\mathbf{m} &= \mathbf{m}^{\setminus i} + \mathbf{V}^{\setminus i}\nabla_m \log Z_i && (7)\\
\mathbf{V} &= \mathbf{V}^{\setminus i} - \mathbf{V}^{\setminus i}(\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i)\mathbf{V}^{\setminus i} && (8)
\end{aligned}$$

(Warning: if the prior distribution is uniform, then $\mathbf{V}^{\setminus i} = \infty$ and you need to take appropriate limits in the formulas above and below.) To update the approximation:

$$\begin{aligned}
\mathbf{V}_i^{-1} &= \mathbf{V}^{-1} - (\mathbf{V}^{\setminus i})^{-1} && (9)\\
\mathbf{V}_i &= (\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i)^{-1} - \mathbf{V}^{\setminus i} && (10)\\
\mathbf{m}_i &= \mathbf{V}_i(\mathbf{V}^{-1}\mathbf{m} - (\mathbf{V}^{\setminus i})^{-1}\mathbf{m}^{\setminus i}) && (11)\\
&= \mathbf{m}^{\setminus i} + (\mathbf{V}_i + \mathbf{V}^{\setminus i})(\mathbf{V}^{\setminus i})^{-1}(\mathbf{m} - \mathbf{m}^{\setminus i}) && (12)\\
&= \mathbf{m}^{\setminus i} + (\mathbf{V}_i + \mathbf{V}^{\setminus i})\nabla_m \log Z_i && (13)\\
&= \mathbf{m}^{\setminus i} + (\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i)^{-1}\nabla_m \log Z_i && (14)\\
s_i &= Z_i \frac{\left|\mathbf{V}_i + \mathbf{V}^{\setminus i}\right|^{1/2}}{\left|\mathbf{V}_i\right|^{1/2}} \exp(\frac{1}{2}(\mathbf{m}_i - \mathbf{m}^{\setminus i})^{\mathrm{T}}(\mathbf{V}_i + \mathbf{V}^{\setminus i})^{-1}(\mathbf{m}_i - \mathbf{m}^{\setminus i})) && (15)\\
&= Z_i \left|\mathbf{I} + \mathbf{V}^{\setminus i}\mathbf{V}_i^{-1}\right|^{1/2} \exp(\frac{1}{2}\nabla_m^{\mathrm{T}}(\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v)^{-1}\nabla_m \log Z_i) && (16)
\end{aligned}$$

1.1 Rank 1 updates

An important special case is when the derivatives have rank one:

$$\begin{aligned}
\nabla_m \log Z_i &= \alpha_i \mathbf{x}_i && (17)\\
\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i &= \beta_i \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} && (18)\\
\mathbf{x}_i^{\mathrm{T}}(\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i)^{-1}\mathbf{x}_i &= \beta_i^{-1} && (19)
\end{aligned}$$

Here (α_i, β_i) are scalars and $\mathbf{x}_i$ is some vector. In this case, you can use a special representation:

$$\mathbf{V}_i^{-1} = v_i^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}} \tag{20}$$
$$\mathbf{x}_i^{\mathrm{T}}\mathbf{V}_i\mathbf{x}_i = v_i \tag{21}$$
$$m_i = \mathbf{x}_i^{\mathrm{T}}\mathbf{m}_i \tag{22}$$

To remove such a term:

$$\mathbf{V}^{\setminus i} = (\mathbf{V}^{-1} - v_i^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}})^{-1} \tag{23}$$
$$= \mathbf{V} + (\mathbf{V}\mathbf{x}_i)(v_i - \mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)^{-1}(\mathbf{x}_i^{\mathrm{T}}\mathbf{V}) \tag{24}$$
$$\mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i = \mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i(1 - v_i^{-1}\mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)^{-1} \tag{25}$$
$$\mathbf{m}^{\setminus i} = \mathbf{m} + \mathbf{V}^{\setminus i}\mathbf{V}_i^{-1}(\mathbf{m} - \mathbf{m}_i) \tag{26}$$
$$= \mathbf{m} + \mathbf{V}^{\setminus i}\mathbf{x}_i v_i^{-1}(\mathbf{x}_i^{\mathrm{T}}\mathbf{m} - m_i) \tag{27}$$
$$= \mathbf{m} + (\mathbf{V}\mathbf{x}_i)(1 - v_i^{-1}\mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)^{-1}v_i^{-1}(\mathbf{x}_i^{\mathrm{T}}\mathbf{m} - m_i) \tag{28}$$
$$\mathbf{x}_i^{\mathrm{T}}\mathbf{m}^{\setminus i} = \mathbf{x}_i^{\mathrm{T}}\mathbf{m} + (\mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)(1 - v_i^{-1}\mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)^{-1}v_i^{-1}(\mathbf{x}_i^{\mathrm{T}}\mathbf{m} - m_i) \tag{29}$$

To update the term:

$$m_i = \mathbf{x}_i^{\mathrm{T}}\mathbf{m}^{\setminus i} + \frac{\alpha_i}{\beta_i} \tag{30}$$
$$v_i = \beta_i^{-1} - \mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i \tag{31}$$
$$s_i = Z_i(\beta_i v_i)^{-1/2}\exp(\frac{\alpha_i^2}{2\beta_i}) \tag{32}$$
$$= Z_i\sqrt{1 + \mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i v_i^{-1}}\exp(\frac{\alpha_i^2}{2\beta_i}) \tag{33}$$

Using these equations, $\mathbf{V}^{\setminus i}$ and $\mathbf{m}^{\setminus i}$ need not be computed in full, but only their projections onto $\mathbf{x}_i$. The new $q(\mathbf{w})$ can be computed directly from the old $q(\mathbf{w})$:

$$d_i = (1 - (v_i^{old})^{-1}\mathbf{x}_i^{\mathrm{T}}\mathbf{V}\mathbf{x}_i)^{-1} \tag{34}$$
$$\mathbf{m}^{new} = \mathbf{m}^{\setminus i} + \mathbf{V}^{\setminus i}\mathbf{x}_i\alpha_i \tag{35}$$
$$= \mathbf{m} + \mathbf{V}\mathbf{x}_i d_i((v_i^{old})^{-1}(\mathbf{x}_i^{\mathrm{T}}\mathbf{m} - m_i^{old}) + \alpha_i) \tag{36}$$
$$\mathbf{V}^{new} = \mathbf{V}^{\setminus i} - (\mathbf{V}^{\setminus i}\mathbf{x}_i)\beta_i(\mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}) \tag{37}$$
$$= \mathbf{V} + (\mathbf{V}\mathbf{x}_i)d_i((v_i^{old})^{-1} - \beta_i d_i)(\mathbf{x}_i^{\mathrm{T}}\mathbf{V}) \tag{38}$$

Alternatively, using natural parameters for the messages:

$$r = \frac{\beta_i}{(\mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i)^{-1} - \beta_i} \tag{39}$$
$$(v_i^{new})^{-1} = r(\mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i)^{-1} \tag{40}$$
$$(v_i^{new})^{-1}m_i^{new} = r\left(\alpha + \frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{m}^{\setminus i}}{\mathbf{x}_i^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}_i}\right) + \alpha \tag{41}$$

$$\Delta(v_i^{-1}) = (v_i^{new})^{-1} - (v_i^{old})^{-1} \tag{42}$$

$$\Delta(v_i^{-1} m_i) = (v_i^{new})^{-1} m_i^{new} - (v_i^{old})^{-1} m_i^{old} \tag{43}$$

$$\mathbf{V}^{new} = \mathbf{V} - (\mathbf{V}\mathbf{x}_i) \frac{\Delta(v_i^{-1})}{1 + \mathbf{x}_i^{\mathrm{T}} \mathbf{V} \mathbf{x}_i \Delta(v_i^{-1})} (\mathbf{x}_i^{\mathrm{T}} \mathbf{V}) \tag{44}$$

$$\mathbf{m}^{new} = \mathbf{m} + (\mathbf{V}\mathbf{x}_i) \frac{\Delta(v_i^{-1} m_i) - \mathbf{x}_i^{\mathrm{T}} \mathbf{m} \Delta(v_i^{-1})}{1 + \mathbf{x}_i^{\mathrm{T}} \mathbf{V} \mathbf{x}_i \Delta(v_i^{-1})} \tag{45}$$

This completes the computation for one term.

If the prior is $p(\mathbf{w}) \sim \mathcal{N}(\mathbf{m}_0, \mathbf{V}_0)$, then the algorithm maintains the invariant that

$$\mathbf{V}^{-1} = \mathbf{V}_0^{-1} + \sum_i v_i^{-1} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \tag{46}$$

$$\mathbf{m} = \mathbf{V}(\mathbf{V}_0^{-1} \mathbf{m}_0 + \sum_i v_i^{-1} m_i \mathbf{x}_i) \tag{47}$$

And the final normalizing constant is

$$s = \frac{|\mathbf{V}|^{1/2}}{|\mathbf{V}_0|^{1/2}} \exp(B/2) \prod_{i=1}^{n} s_i \tag{48}$$

$$\text{where } B = \mathbf{m}^{\mathrm{T}} \mathbf{V}^{-1} \mathbf{m} - \mathbf{m}_0^{\mathrm{T}} \mathbf{V}_0^{-1} \mathbf{m}_0 - \sum_i \frac{m_i^2}{v_i} \tag{49}$$

1.2 Example: step function

This has the special form.

$$f_i(\mathbf{w}) = \epsilon + (1 - 2\epsilon)\Theta(\mathbf{x}^{\mathrm{T}} \mathbf{w}) \tag{50}$$

$$\phi(z) = \int_{-\infty}^{z} \mathcal{N}(z; 0, 1) dz \tag{51}$$

$$z = \frac{\mathbf{x}^{\mathrm{T}} \mathbf{m}^{\setminus i}}{\sqrt{\mathbf{x}^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{x}}} \tag{52}$$

$$Z(\mathbf{m}^{\setminus i}, \mathbf{V}^{\setminus i}) = \epsilon + (1 - 2\epsilon)\phi(z) \tag{53}$$

$$\alpha = \frac{1}{\sqrt{\mathbf{x}^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{x}}} \frac{(1 - 2\epsilon)\mathcal{N}(z; 0, 1)}{\epsilon + (1 - 2\epsilon)\phi(z)} \tag{54}$$

$$\beta = \alpha \left(\alpha + \frac{\mathbf{x}^{\mathrm{T}} \mathbf{m}^{\setminus i}}{\mathbf{x}^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{x}} \right) \tag{55}$$

1.3 Example: probit function

This has the special form.

$$
\begin{aligned}
f_i(\mathbf{w}) &= \phi(\mathbf{x}^{\mathrm{T}}\mathbf{w}) && (56)\\
z &= \frac{\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i}}{\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}+1}} && (57)\\
Z(\mathbf{m}^{\setminus i},\mathbf{V}^{\setminus i}) &= \phi(z) && (58)\\
\alpha &= \frac{1}{\sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}+1}}\frac{\mathcal{N}(z;0,1)}{\phi(z)} && (59)\\
\beta &= \alpha\left(\alpha+\frac{\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i}}{\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}+1}\right) && (60)
\end{aligned}
$$

1.4 Example: logistic function

$$
f_i(\mathbf{w}) = \sigma(\mathbf{x}^{\mathrm{T}}\mathbf{w}) = \frac{1}{1+\exp(-\mathbf{x}^{\mathrm{T}}\mathbf{w})} \quad (61)
$$

This cannot be handled exactly, but there is a handy approximation, due to MacKay (1992):

$$
\begin{aligned}
z &= \frac{\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i}}{\sqrt{1+(\pi/8)\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}}} && (62)\\
Z(\mathbf{m}^{\setminus i},\mathbf{V}^{\setminus i}) &\approx \sigma(z) && (63)\\
\alpha &= \frac{\sigma(-z)}{\sqrt{1+(\pi/8)\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}}} && (64)\\
\beta &= \alpha\frac{\alpha+(\pi/8)\mathbf{x}^{\mathrm{T}}\mathbf{m}}{1+(\pi/8)\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}} && (65)
\end{aligned}
$$

Another approach is Gauss-Hermite quadrature. Let (z_k, c_k) be the nodes and weights for integration against $\mathcal{N}(\mathbf{x}^{\mathrm{T}}\mathbf{m}, \mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x})$.

$$
\begin{aligned}
Z(\mathbf{m}^{\setminus i},\mathbf{V}^{\setminus i}) &\approx \sum_k c_k\sigma(z_k)\frac{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i},\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x})}{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m},\mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x})} && (66)\\
Z_1 &= \sum_k c_k z_k\sigma(z_k)\frac{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i},\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x})}{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m},\mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x})} && (67)\\
Z_2 &= \sum_k c_k z_k^2\sigma(z_k)\frac{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i},\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x})}{\mathcal{N}(z_k;\mathbf{x}^{\mathrm{T}}\mathbf{m},\mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x})} && (68)\\
\alpha &= \frac{1}{\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}}\left(\frac{Z_1}{Z}-\mathbf{x}^{\mathrm{T}}\mathbf{m}^{\setminus i}\right) && (69)\\
\beta &= \frac{1}{\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x}}-\frac{1}{(\mathbf{x}^{\mathrm{T}}\mathbf{V}^{\setminus i}\mathbf{x})^2}\left(\frac{Z_2}{Z}-\left(\frac{Z_1}{Z}\right)^2\right) && (70)
\end{aligned}
$$

2 GP EP

Take the simplified equations in the previous section and rewrite them in terms of inner products. Define

$$\lambda_i = \mathbf{x}_i^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{x}_i \tag{71}$$

$$C_{ij} = \mathbf{x}_i^{\mathrm{T}} \mathbf{x}_j \tag{72}$$

$$\mathbf{\Lambda} = \mathrm{diag}(v_1, ..., v_n) \tag{73}$$

$$\mathbf{A} = (\mathbf{C}^{-1} + \mathbf{\Lambda}^{-1})^{-1} \tag{74}$$

$$h_i = \mathbf{x}_i^{\mathrm{T}} \mathbf{m} \tag{75}$$

$$h_i^{\setminus i} = \mathbf{x}_i^{\mathrm{T}} \mathbf{m}^{\setminus i} \tag{76}$$

Deletion step:

$$h_i = \sum_j A_{ij} \frac{m_j}{v_j} \tag{77}$$

$$h_i^{\setminus i} = h_i + \lambda_i v_i^{-1} (h_i - m_i) \tag{78}$$

$$\lambda_i = \frac{a_{ii}}{1 - a_{ii} v_i^{-1}} \tag{79}$$

First part of ADF:

$$h_i = h_i^{\setminus i} + \lambda_i \alpha_i \tag{80}$$

Update:

$$v_i = \beta_i^{-1} - \lambda_i \tag{81}$$

$$m_i = h_i^{\setminus i} + \frac{\alpha_i}{\beta_i} \tag{82}$$

$$v_i^{-1} m_i = v_i^{-1} h_i + \alpha_i \qquad \text{(natural parameter)} \tag{83}$$

Second part of ADF:

$$\mathbf{A} = \mathbf{A}^{old} - \frac{\mathbf{a}_i \mathbf{a}_i^{\mathrm{T}}}{\delta + a_{ii}} \tag{84}$$

$$\text{where } \delta = \left(\frac{1}{v_i^{new}} - \frac{1}{v_i^{old}} \right)^{-1} \tag{85}$$

This completes the computation for one term. Seeger (2002) suggests choosing terms i based on the differential entropy score:

$$\Delta \mathcal{H}(q(\mathbf{w})) = \frac{1}{2} \log \frac{\lambda_i^{new}}{a_{ii}} = \frac{1}{2} \log \frac{1 - \lambda_i \beta_i}{1 - a_{ii} v_i^{-1}} \tag{86}$$

Including a standard prior $p(\mathbf{w}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ gives the final normalizing constant:

$$B = \sum_{ij} A_{ij} \frac{m_i m_j}{v_i v_j} - \sum_i \frac{m_i^2}{v_i} \tag{87}$$

$$p(D) = \frac{|\Lambda|^{1/2}}{|\mathbf{C} + \Lambda|^{1/2}} \exp(B/2) \prod_{i=1}^{n} s_i \tag{88}$$

2.1 Example: step function

$$
\begin{aligned}
z &= \frac{h_i^{\setminus i}}{\sqrt{\lambda_i}} && (89) \\
\alpha &= \frac{1}{\sqrt{\lambda_i}} \frac{(1-2\epsilon)\mathcal{N}(z;0,1)}{\epsilon + (1-2\epsilon)\phi(z)} && (90) \\
\beta &= \alpha \left(\alpha + \frac{h_i^{\setminus i}}{\lambda_i} \right) = \alpha \frac{h_i}{\lambda_i} && (91)
\end{aligned}
$$

2.2 Example: probit function

$$
\begin{aligned}
z &= \frac{h_i^{\setminus i}}{\sqrt{\lambda_i + 1}} && (92) \\
\alpha &= \frac{1}{\sqrt{\lambda_i + 1}} \frac{\mathcal{N}(z;0,1)}{\phi(z)} && (93) \\
\beta &= \alpha \left(\alpha + \frac{h_i^{\setminus i}}{\lambda_i + 1} \right) = \alpha \frac{h_i + \alpha}{\lambda_i + 1} && (94)
\end{aligned}
$$

3 Rank k updates

The techniques for rank 1 update can be generalized to rank k. Suppose the derivatives are:

$$
\begin{aligned}
\nabla_m \log Z_i &= \mathbf{X}_i \mathbf{a}_i && (95) \\
\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i &= \mathbf{X}_i \mathbf{B}_i \mathbf{X}_i^{\mathrm{T}} && (96) \\
\mathbf{X}_i^{\mathrm{T}} (\nabla_m \nabla_m^{\mathrm{T}} - 2\nabla_v \log Z_i)^{-1} \mathbf{X}_i &= \mathbf{B}_i^{-1} && (97)
\end{aligned}
$$

Here $\mathbf{X}_i$ is a matrix with k columns and any number of rows, $\mathbf{a}_i$ is a k by 1 vector, and $\mathbf{B}_i$ is a k by k matrix. The big $(\mathbf{V}_i, \mathbf{m}_i)$ can now be represented by a small $k \times k$ matrix v_i and $k \times 1$ vector m_i:

$$
\begin{aligned}
\mathbf{V}_i^{-1} &= \mathbf{X}_i v_i^{-1} \mathbf{X}_i^{\mathrm{T}} && (98) \\
\mathbf{X}_i^{\mathrm{T}} \mathbf{V}_i \mathbf{X}_i &= v_i && (99) \\
m_i &= \mathbf{X}_i^{\mathrm{T}} \mathbf{m}_i && (100)
\end{aligned}
$$

To remove such a term:

$$
\begin{aligned}
\mathbf{V}^{\setminus i} &= (\mathbf{V}^{-1} - \mathbf{X}_i v_i^{-1} \mathbf{X}_i^{\mathrm{T}})^{-1} && (101) \\
&= \mathbf{V} + (\mathbf{V}\mathbf{X}_i)(v_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i)^{-1} (\mathbf{X}_i^{\mathrm{T}} \mathbf{V}) && (102) \\
\mathbf{X}_i^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{X}_i &= \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i (v_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i)^{-1} v_i && (103) \\
\mathbf{m}^{\setminus i} &= \mathbf{m} + \mathbf{V}^{\setminus i} \mathbf{V}_i^{-1} (\mathbf{m} - \mathbf{m}_i) && (104) \\
&= \mathbf{m} + \mathbf{V}^{\setminus i} \mathbf{X}_i v_i^{-1} (\mathbf{X}_i^{\mathrm{T}} \mathbf{m} - m_i) && (105) \\
&= \mathbf{m} + (\mathbf{V}\mathbf{X}_i)(v_i - \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i)^{-1} (\mathbf{X}_i^{\mathrm{T}} \mathbf{m} - m_i) && (106)
\end{aligned}
$$

To update the term:

$$
\begin{aligned}
m_i &= \mathbf{X}_i^{\mathrm{T}} \mathbf{m}^{\setminus i} + \mathbf{B}_i^{-1} \mathbf{a}_i & (107) \\
v_i &= \mathbf{B}_i^{-1} - \mathbf{X}_i^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{X}_i & (108) \\
s_i &= Z_i \left| \mathbf{B}_i v_i \right|^{-1/2} \exp(\frac{1}{2} \mathbf{a}_i^{\mathrm{T}} \mathbf{B}_i^{-1} \mathbf{a}_i) & (109) \\
&= Z_i \left| 1 + \mathbf{X}_i^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{X}_i v_i^{-1} \right|^{1/2} \exp(\frac{1}{2} \mathbf{a}_i^{\mathrm{T}} \mathbf{B}_i^{-1} \mathbf{a}_i) & (110)
\end{aligned}
$$

Using these equations, $\mathbf{V}^{\setminus i}$ and $\mathbf{m}^{\setminus i}$ need not be computed in full, but only their projections onto $\mathbf{X}_i$. The new $q(\mathbf{w})$ can be computed directly from the old $q(\mathbf{w})$:

$$
\begin{aligned}
d_i &= (1 - (v_i^{old})^{-1} \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i)^{-1} & (111) \\
d_i' &= (1 - \mathbf{X}_i^{\mathrm{T}} \mathbf{V} \mathbf{X}_i (v_i^{old})^{-1})^{-1} = \mathbf{I} + \mathbf{X}_i^{\mathrm{T}} \mathbf{V}^{\setminus i} \mathbf{X}_i (v_i^{old})^{-1} & (112) \\
\mathbf{m}^{new} &= \mathbf{m}^{\setminus i} + \mathbf{V}^{\setminus i} \mathbf{X}_i \mathbf{a}_i & (113) \\
&= \mathbf{m} + \mathbf{V} \mathbf{X}_i d_i ((v_i^{old})^{-1} (\mathbf{X}_i^{\mathrm{T}} \mathbf{m} - m_i^{old}) + \mathbf{a}_i) & (114) \\
\mathbf{V}^{new} &= \mathbf{V}^{\setminus i} - (\mathbf{V}^{\setminus i} \mathbf{X}_i) \mathbf{B}_i (\mathbf{X}_i^{\mathrm{T}} \mathbf{V}^{\setminus i}) & (115) \\
&= \mathbf{V} + (\mathbf{V} \mathbf{X}_i) d_i ((v_i^{old})^{-1} - \mathbf{B}_i d_i') (\mathbf{X}_i^{\mathrm{T}} \mathbf{V}) & (116)
\end{aligned}
$$

Acknowledgement

Thanks to Ralf Herbrich for optimizing the rank 1 and rank k updates.

References

MacKay, D. J. (1992). The evidence framework applied to classification networks. *Neural Computation*, *4*, 698–714. `http://www.inference.phy.cam.ac.uk/mackay/PhD.html`.

Seeger, M. (2002). Notes on Minka's Expectation Propagation for Gaussian Process classification (Technical Report). University of Edinburgh. `http://www.cs.berkeley.edu/~mseeger/papers/ivm.html#minkanotes`.