


Appendix

Throughout the appendix the expectations and variances of random variables are taken with respect to a Gaussian distribution with parameters $\mathbf{C}, \mu$ (or $\tilde{C}, \tilde{\mu}$). We list several algebraic properties that are used in the paper to derive the updates.

1. $\text{tr}(\mathbf{A}\mathbf{B}) = \sum_{p=1}^{n} \sum_{q=1}^{m} A_{pq} B_{qp} = \sum_{q=1}^{m} \sum_{p=1}^{n} B_{qp} A_{pq} = \text{tr} (\mathbf{B} \mathbf{A})$.

2. Let $A$ be a square matrix. Then, $|A| = \sum_p \text{cof}_{pq} A_{pq}$ and $A^{-1} = \frac{\text{cof}_{pq}}{|A|}$, where $\text{cof}_{pq}$ is the cofactor of $A_{pq}$ in $A$. For a symmetric matrix $A$, $\text{cof}_{pq} = \text{cof}_{qp}$ and we get that

$$\frac{d |A|}{d A_{pq}} = \text{cof}_{p,q} = \text{cof}_{q,p} = |A|(A^{-1})_{p,q}.$$ 

Thus $\frac{d |A|}{d A} = |A| A^{-1}$ and, similarly, $\frac{d A^{-1}}{d A} = -|A|^{-1} A^{-1}$. 

11
3. From (2) we get \( \frac{\partial \ln |A|}{\partial A} = A^{-1} \).

4. From (2) and (3) we have \( |A^{-1}| = |A|^{-1} \) and thus \( \frac{\partial \ln |A^{-1}|}{\partial A^{-1}} = -\frac{\partial \ln |A|}{\partial A^{-1}} = -A \).

5. 
\[
C = E((x - \mu)(x - \mu)^T) = E(xx^T) - E(x\mu^T) - E(\mu x^T) + E(\mu \mu^T) \\
= E(xx^T) - E(\mu \mu^T)
\]

6. 
\[
\tilde{E}((x - \tilde{\mu})^T \tilde{C}^{-1}(x - \tilde{\mu})) = \tilde{E}(\text{tr}( (x - \tilde{\mu})^T \tilde{C}^{-1}(x - \tilde{\mu}) )) \\
= \tilde{E}(\text{tr}( \tilde{C}^{-1}(x - \tilde{\mu})(x - \tilde{\mu})^T )) \\
= \text{tr}(\tilde{C}^{-1} \tilde{E}(x - \tilde{\mu})(x - \tilde{\mu})^T) \\
= \text{tr}(\tilde{C}^{-1} \tilde{C}) = \text{tr}(I) = d
\]

7. 
\[
\tilde{E}((x - \mu)^T C^{-1}(x - \mu)) = \tilde{E}(\text{tr}( (x - \mu)^T C^{-1}(x - \mu) )) \\
= \text{tr}(C^{-1} \tilde{E}(x - \mu)(x - \mu)^T) \\
= \text{tr}(C^{-1} \tilde{E}(xx^T - x\mu^T - \mu x^T + \mu \mu^T)) \\
= \text{tr}(C^{-1}(C + \tilde{\mu} \tilde{\mu}^T - \tilde{\mu} \mu^T - \mu \tilde{\mu}^T + \mu \mu^T)) \\
= \text{tr}(C^{-1} \tilde{C}) + \text{tr}(C^{-1}(\tilde{\mu} - \mu)(\tilde{\mu} - \mu)^T) \\
= \text{tr}(C^{-1} \tilde{C}) + (\tilde{\mu} - \mu)^T C^{-1}(\tilde{\mu} - \mu)
\]

8. 
\[
\frac{\partial \text{tr}(AB)}{\partial B} = \sum_p \sum_q A_{pq} B_{qp} = A^T
\]

9. Let \( I_{pq} \) denotes the matrix that has a one in position \((p,q)\) and is zero otherwise and let \( A_{*,p}(A_{p,*}) \) denote the \( p \)th column (row) of the matrix \( A \). Then,
\[
AI_{pq} B = \text{col}(A, p) \cdot \text{row}(B, q) = A_{*,p} B_{q,*} \quad \text{and} \quad x^T I_{pq} x = [x^T x]_{pq} = x_p x_q
\]

10. Let \( 0 \) denote the zero matrix. Then,
\[
0 = \frac{dI}{d C_{pq}} = \frac{dC^{-1}C}{d C_{pq}} = \frac{dC^{-1}}{d C_{pq}} C + C^{-1} \frac{dC}{d C_{pq}}
\]
This implies that \( \frac{dC^{-1}}{d C_{pq}} = -C^{-1} I_{pq} C^{-1} = -C^{-1}_{pq} C^{-1}_{qp} \).

11. \( \frac{\partial C^{-1}}{\partial C} = -C^{-1} C^{-1} \).

12. \( \frac{\partial x^T A x}{\partial A} = \frac{\partial (xx^T A)}{\partial A} = (xx^T)^T = x x^T \)

13. \( \frac{d x^T B x}{d y} = x^T \frac{d B}{d y} x \)

14. From above we get,
\[
\frac{d(x - \mu)^T C^{-1}(x - \mu)}{d C_{pq}} = -(x - \mu)^T C^{-1} I_{pq} C^{-1}(x - \mu) \\
= -(C^{-1}(x - \mu))^T I_{pq} C^{-1}(x - \mu) \\
= -(C^{-1}(x - \mu))_p (C^{-1}(x - \mu))_q
\]
Hence
\[
\frac{\partial (x - \mu)^T C^{-1} (x - \mu)}{\partial C} = -(C^{-1}(x - \mu))(C^{-1}(x - \mu))^T = -C^{-1}(x - \mu)(x - \mu)^T C^{-1}
\]

15. Equality (14) imply that \( \frac{\partial (x - \mu)^T C^{-1}(x - \mu)}{\partial C} = (x - \mu)(x - \mu)^T. \)

16. Let A and B be symmetric matrices. Then,
\[
\frac{\partial}{\partial B_{ij}} \sum_{pq} A_{pq} B_{pq} = \sum_{pq} A_{pq} \frac{\partial B_{pq}}{\partial B_{ij}} = -\sum_{pq} A_{pq} B_{ip} B_{pq} = -B_{ij} A B_{ij}
\]

Hence, \( \frac{\partial \text{tr}(AB)}{\partial B^{-1}} = -B A B. \)