Lecture Notes: Gaussian identities

Marc Toussaint
Machine Learning & Robotics group, TU Berlin
Franklinstr. 28/29, FR 6-9, 10587 Berlin, Germany

October 14, 2008

Definitions
We define a Gaussian over x with mean a and covariance matrix A as the function

$$\mathcal{N}(x \,|\, a, A) = \frac{1}{|2\pi A|^{1/2}}\ \exp\{-\frac{1}{2}(x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\} \tag{1}$$

with property $N(x \,|\, a, A) = N(a|\ x, A)$. We also define the canonical form with precision matrix A as

$$\mathcal{N}[x \,|\, a, A] = \frac{\exp\{-\frac{1}{2}a^\top A^{\text{-}1} a\}}{|2\pi A^{\text{-}1}|^{1/2}}\ \exp\{-\frac{1}{2}x^\top\ A\ x + x^\top a\} \tag{2}$$

with properties

$$\mathcal{N}[x \,|\, a, A] = \mathcal{N}(x \,|\, A^{\text{-}1}a, A^{\text{-}1}) \tag{3}$$
$$\mathcal{N}(x \,|\, a, A) = \mathcal{N}[x \,|\, A^{\text{-}1}a, A^{\text{-}1}]\ . \tag{4}$$

Non-normalized Gaussian

$$\overline{\mathcal{N}}(x, a, A) = |2\pi A|^{1/2}\ \mathcal{N}(x|a, A) \tag{5}$$
$$= \exp\{-\frac{1}{2}(x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\} \tag{6}$$

Matrices [matrix cookbook]

$$(A^{\text{-}1} + B^{\text{-}1})^{\text{-}1} = A\ (A\text{+}B)^{\text{-}1}\ B = B\ (A\text{+}B)^{\text{-}1}\ A \tag{7}$$
$$(A^{\text{-}1} - B^{\text{-}1})^{\text{-}1} = A\ (B\text{-}A)^{\text{-}1}\ B \tag{8}$$
$$\partial_x|A_x| = |A_x|\ \mathrm{tr}(A_x^{\text{-}1}\ \partial_x A_x) \tag{9}$$
$$\partial_x A_x^{\text{-}1} = -A_x^{\text{-}1}\ (\partial_x A_x)\ A_x^{\text{-}1} \tag{10}$$
$$(A + UBV)^{\text{-}1} = A^{\text{-}1} - A^{\text{-}1}U(B^{\text{-}1} + VA^{\text{-}1}U)^{\text{-}1}VA^{\text{-}1} \tag{11}$$
$$(A^{\text{-}1} + B^{\text{-}1})^{\text{-}1} = A - A(B + A)^{\text{-}1}A \tag{12}$$
$$(A + J^\top BJ)^{\text{-}1}J^\top B = A^{\text{-}1}J^\top(B^{\text{-}1} + JA^{\text{-}1}J^\top)^{\text{-}1} \tag{13}$$
$$(A + J^\top BJ)^{\text{-}1}A = \mathbf{I} - (A + J^\top BJ)^{\text{-}1}J^\top BJ \tag{14}$$

(11)=Woodbury; (13,14) holds for pos def A and B

Derivatives

$$\partial_x\mathcal{N}(x|a, A) = \mathcal{N}(x|a, A)\ (-h^\top)\ , \quad h := A^{\text{-}1}(x\text{-}a) \tag{15}$$
$$\partial_\theta\mathcal{N}(x|a, A) = \mathcal{N}(x|a, A)\ \cdot$$
$$\Big[-h^\top(\partial_\theta x) + h^\top(\partial_\theta a) - \frac{1}{2}\mathrm{tr}(A^{\text{-}1}\ \partial_\theta A) + \frac{1}{2}h^\top(\partial_\theta A)h\Big] \tag{16}$$
$$\partial_\theta\mathcal{N}[x|a, A] = \mathcal{N}[x|a, A]\ \Big[-\frac{1}{2}x^\top\partial_\theta Ax + \frac{1}{2}a^\top A^{\text{-}1}\partial_\theta AA^{\text{-}1}a$$
$$+ x^\top\partial_\theta a - a^\top A^{\text{-}1}\partial_\theta a + \frac{1}{2}\mathrm{tr}(\partial_\theta AA^{\text{-}1})\Big] \tag{17}$$
$$\partial_\theta\overline{\mathcal{N}}_x(a, A) = \overline{\mathcal{N}}_x(a, A)\ \cdot$$
$$\Big[h^\top(\partial_\theta x) + h^\top(\partial_\theta a) + \frac{1}{2}h^\top(\partial_\theta A)h\Big] \tag{18}$$

Product
The product of two Gaussians can be expressed as

$$\mathcal{N}(x \,|\, a, A)\ \mathcal{N}(x \,|\, b, B)$$
$$= \mathcal{N}[x \,|\, A^{\text{-}1}a + B^{\text{-}1}b, A^{\text{-}1} + B^{\text{-}1}]\ \mathcal{N}(a \,|\, b, A + B)\ , \tag{19}$$
$$\mathcal{N}[x \,|\, a, A]\ \mathcal{N}[x \,|\, b, B]$$
$$= \mathcal{N}[x \,|\, a + b, A + B]\ \mathcal{N}(A^{\text{-}1}a \,|\, B^{\text{-}1}b, A^{\text{-}1} + B^{\text{-}1}) \tag{20}$$
$$= \mathcal{N}[x \,|\, a + b, A + B]$$
$$\mathcal{N}[A^{\text{-}1}a \,|\, A(A\text{+}B)^{\text{-}1}b, A(A\text{+}B)^{\text{-}1}B] \tag{21}$$
$$= \mathcal{N}[x \,|\, a + b, A + B]$$
$$\mathcal{N}[A^{\text{-}1}a \,|\, b - B(A\text{+}B)^{\text{-}1}b, B - B(A\text{+}B)^{\text{-}1}B]\ , \tag{22}$$
$$\mathcal{N}(x \,|\, a, A)\ \mathcal{N}[x \,|\, b, B]$$
$$= \mathcal{N}[x \,|\, A^{\text{-}1}a + b, A^{\text{-}1} + B]\ \mathcal{N}(a \,|\, B^{\text{-}1}b, A + B^{\text{-}1})\ . \tag{23}$$

Convolution

$$\int_x \mathcal{N}(x \,|\, a, A)\ \mathcal{N}(y - x \,|\, b, B)\ dx = \mathcal{N}(y \,|\, a + b, A + B) \tag{24}$$

Division

$$\mathcal{N}(x|a, A)\ /\ \mathcal{N}(x|b, B) = \mathcal{N}(x|c, C)\ /\ \mathcal{N}(c|b, C + B)$$
$$C^{\text{-}1}c = A^{\text{-}1}a - B^{\text{-}1}b$$
$$C^{\text{-}1} = A^{\text{-}1} - B^{\text{-}1} \tag{25}$$
$$\mathcal{N}[x|a, A]\ /\ \mathcal{N}[x|b, B] \propto \mathcal{N}[x|a - b, A - B] \tag{26}$$

Expectations
Let $x \sim \mathcal{N}(x \,|\, a, A)$,

$$\mathrm{E}_x\{g(x)\} := \int_x \mathcal{N}(x \,|\, a, A)\ g(x)\ dx \tag{27}$$
$$\mathrm{E}_x\{x\} = a\ , \quad \mathrm{E}_x\{xx^\top\} = A \tag{28}$$
$$\mathrm{E}_x\{f + Fx\} = f + Fa \tag{29}$$
$$\mathrm{E}_x\{x^\top x\} = a^\top a + \mathrm{tr}(A) \tag{30}$$
$$\mathrm{E}_x\{(x\text{-}m)^\top R(x\text{-}m)\} = (a\text{-}m)^\top R(a\text{-}m) + \mathrm{tr}(RA) \tag{31}$$

Transformation
Linear transformations in x imply the following identities,

$$\mathcal{N}(Fx+f\,|\,a,A) = \frac{1}{|F|}\,\mathcal{N}(x\,|\ F^{\text{-}1}(a-f),\ F^{\text{-}1}AF^{\text{-}\top}) = \frac{1}{|F|}\,\mathcal{N}[x\,|\ F^\top A^{\text{-}1}(a-f),\ F^\top A^{\text{-}1}F]\ , \tag{32}$$

$$\mathcal{N}[Fx+f\,|\,a,A] = \frac{1}{|F|}\,\mathcal{N}[x\,|\ F^\top(a-Af),\ F^\top AF]\ . \tag{33}$$

marginal & conditional:

$$\mathcal{N}(x\,|\,a,A)\ \mathcal{N}(y\,|\,b+Fx,B) = \mathcal{N}\big(\begin{pmatrix}x\\y\end{pmatrix}\big|\begin{pmatrix}a\\b+Fa\end{pmatrix},\ \begin{pmatrix}A & A^\top F^\top\\ FA & B{+}FA^\top F^\top\end{pmatrix}\big) \tag{34}$$

$$\mathcal{N}\big(\begin{pmatrix}x\\y\end{pmatrix}\big|\begin{pmatrix}a\\b\end{pmatrix},\ \begin{pmatrix}A & C\\ C^\top & B\end{pmatrix}\big) = \mathcal{N}(x\,|\,a,A)\cdot\mathcal{N}(y\,|\,b+C^\top A^{\text{-}1}(x\text{-}a),\ B-C^\top A^{\text{-}1}C) \tag{35}$$

$$\mathcal{N}[x\,|\,a,A]\ \mathcal{N}(y\,|\,b+Fx,B) = \mathcal{N}\big[\begin{pmatrix}x\\y\end{pmatrix}\big|\begin{pmatrix}a+F^\top B^{\text{-}1}b\\ B^{\text{-}1}b\end{pmatrix},\ \begin{pmatrix}A+F^\top B^{\text{-}1}F & -F^\top B^{\text{-}1}\\ -B^{\text{-}1}F & B^{\text{-}1}\end{pmatrix}\big] \tag{36}$$

$$\mathcal{N}[x\,|\,a,A]\ \mathcal{N}[y\,|\,b+Fx,B] = \mathcal{N}\big[\begin{pmatrix}x\\y\end{pmatrix}\big|\begin{pmatrix}a+F^\top B^{\text{-}1}b\\ b\end{pmatrix},\ \begin{pmatrix}A+F^\top B^{\text{-}1}F & -F^\top\\ -F & B\end{pmatrix}\big] \tag{37}$$

$$\mathcal{N}\big[\begin{pmatrix}x\\y\end{pmatrix}\big|\begin{pmatrix}a\\b\end{pmatrix},\ \begin{pmatrix}A & C\\ C^\top & B\end{pmatrix}\big] = \mathcal{N}[x\,|\,a-CB^{\text{-}1}b,\ A-CB^{\text{-}1}C^\top]\cdot\mathcal{N}[y\,|\,b-C^\top x,B] \tag{38}$$

$$\begin{vmatrix}A & C\\ D & B\end{vmatrix} = |A|\ |\widehat{B}| = |\widehat{A}|\ |B|\ \text{, where }\ \begin{array}{l}\widehat{A} = A-CB^{\text{-}1}D\\ \widehat{B} = B-DA^{\text{-}1}C\end{array} \tag{39}$$

$$\begin{bmatrix}A & C\\ D & B\end{bmatrix}^{\text{-}1} = \begin{bmatrix}\widehat{A}^{\text{-}1} & -A^{\text{-}1}C\widehat{B}^{\text{-}1}\\ -\widehat{B}^{\text{-}1}DA^{\text{-}1} & \widehat{B}^{\text{-}1}\end{bmatrix} = \begin{bmatrix}\widehat{A}^{\text{-}1} & -\widehat{A}^{\text{-}1}CB^{\text{-}1}\\ -B^{\text{-}1}D\widehat{A}^{\text{-}1} & \widehat{B}^{\text{-}1}\end{bmatrix} \tag{40}$$

Entropy

$$H(\mathcal{N}(a,A)) = \frac{1}{2}\log|2\pi eA| \tag{41}$$

Kullback-Leibler divergence

$$p=\mathcal{N}(x|a,A)\ ,\quad q=\mathcal{N}(x|b,B)\ ,\quad n=\text{dim}(x)\ ,\quad D\big(p\,\big\|\,q\big) = \sum_x p(x)\log\frac{p(x)}{q(x)} \tag{42}$$

$$2\,D\big(p\,\big\|\,q\big) = \log\frac{|B|}{|A|} + \text{tr}(B^{\text{-}1}A) + (b-a)^\top B^{\text{-}1}(b-a) - n \tag{43}$$

$$4\,D_{\text{sym}}\big(p\,\big\|\,q\big) = \text{tr}(B^{\text{-}1}A) + \text{tr}(A^{\text{-}1}B) + (b-a)^\top(A^{\text{-}1}+B^{\text{-}1})(b-a) - 2n \tag{44}$$

λ-divergence

$$2\,D_\lambda\big(p\,\big\|\,q\big) = \lambda\,D\big(p\,\big\|\,\lambda p+(1{-}\lambda)q\big)\ +\ (1{-}\lambda)\,D\big(p\,\big\|\,(1{-}\lambda)p+\lambda q\big) \tag{45}$$

For $\lambda=.5$: Jensen-Shannon divergence.

Log-likelihoods

$$\log\mathcal{N}(x|a,A) = -\frac{1}{2}\Big[log|2\pi A| + (x\text{-}a)^\top\ A^{\text{-}1}\ (x\text{-}a)\Big] \tag{46}$$

$$\log\mathcal{N}[x|a,A] = -\frac{1}{2}\Big[log|2\pi A^{\text{-}1}| + a^\top A^{\text{-}1}a + x^\top Ax - 2x^\top a\Big] \tag{47}$$

$$\sum_x\mathcal{N}(x|b,B)\log\mathcal{N}(x|a,A) = -D\big(\mathcal{N}(b,B)\,\big\|\,\mathcal{N}(a,A)\big) - H(\mathcal{N}(b,B)) \tag{48}$$

Mixture of Gaussians
Collapsing a MoG into a single Gaussian

$$\underset{b,B}{\text{argmin}}\,D\big(\sum_i p_i\,\mathcal{N}(a_i,A_i)\,\big\|\,\mathcal{N}(b,B)\big)\quad=\quad\Big(b=\sum_i p_ia_i\ ,\ B=\sum_i p_i(A_i+a_ia_i^\top-b\,b^\top)\Big) \tag{49}$$

Marginal of a MOG

$$P(x,y) = \sum_i p_i\ \mathcal{N}(\begin{bmatrix} x \\ y \end{bmatrix} \mid \begin{bmatrix} a_i \\ b_i \end{bmatrix}, \begin{bmatrix} A_i & C_i \\ C_i^\top & B_i \end{bmatrix})$$

$$P(y|x) = \sum_i p_i\ \mathcal{N}(y|b_i + C_i^\top A_i^{\text{-}1}(x - a_i),\ B_i - C_i^\top A_i^{\text{-}1} C_i^\top) \tag{50}$$

$$\approx \mathcal{N}(y|e, E)\ , \quad e = \sum_i p_i(b_i + C_i^\top A_i^{\text{-}1}(x - a_i))\ ,$$

$$E = \sum_i p_i \Big[B_i - C_i^\top A_i^{\text{-}1} C_i^\top + b_i b_i^\top + C_i^\top A_i^{\text{-}1}(x - a_i)(x - a_i)^\top A_i^{\text{-}1\top} C_i + 2C_i^\top A_i^{\text{-}1}(x - a_i) b^\top - e e^\top\Big] \tag{51}$$

$$F = -\sum_i p_i\ C_i^\top A_i^{\text{-}1}\ , \quad f = \sum_i p_i\ (b_i - C_i^\top A_i^{\text{-}1} a_i)\ , \quad Q = ? \tag{52}$$

$$P(x,y) = \sum_i p_i\ \mathcal{N}[\begin{bmatrix} x \\ y \end{bmatrix} \mid \begin{bmatrix} a_i \\ b_i \end{bmatrix}, \begin{bmatrix} A_i & C_i \\ C_i^\top & B_i \end{bmatrix}]$$

$$P(y|x) = \sum_i p_i\ \mathcal{N}[y|b_i - C_i^\top x,\ B_i] \tag{53}$$

$$\approx \mathcal{N}(y|e, E)\ , \quad E = \sum_i p_i(B_i^{\text{-}1} + B^{\text{-}1}(b_i - C_i^\top x)(b_i - C_i^\top x)^\top B^{\text{-}1\top} - e\,e^\top)\ ,$$

$$e = \sum_i p_i\ \Big[B_i^{\text{-}1}(b_i - C_i^\top x)\Big] \tag{54}$$

$$F = -\sum_i p_i\ B_i^{\text{-}1} C_i^\top\ , \quad f = \sum_i p_i\ B_i^{\text{-}1} b_i\ , \quad Q = ? \tag{55}$$

[[todo: unscented transform]