# Additional Exercises on Mathematics

*Latest revision: February 21, 2014*

## Contents



## 1 Fundamentals

**Exercise 1.**
Reduce the following expressions:

*a*) $f(x) = \exp(ax)\exp(b)$

*b*) $f(x) = \exp(ax^2)\exp(bx)\exp(c)$

*c*) $f(x, y) = \exp(ax^5)\exp(by^{-3})\exp(7)$

*d*) $g(z) = (\exp(4z))^2\exp(-3)$

*e*) $h(t) = 1/\exp(2t + 1)$

*f*) $h(t) = \exp(at)/\exp(b)$

**Exercise 2.**
Reduce the following expressions:

*a*) $f(x) = \ln(ax) + \ln(b)$

*b*) $f(x) = 2\ln(\sqrt{a}x) + \ln(b)$

*c*) $f(t) = t\ln(a) + b\ln(a)$

*d*) $g(x) = \ln(x) - \ln(2)$

*e*) $g(y) = \ln(\exp(y))$

*f*) $h(t) = \ln(1)$

**Exercise 3.**
Find the magnitude (length) of the vector $\mathbf{a} = (1, 4, 5)^\top$.

**Exercise 4.**
Given two vectors $\mathbf{a} = (a, b, c)^\top$ and $b = (d, e, f)^\top$, find the matrix $\mathbf{M} = \mathbf{ab}^\top$. What is the rank of $\mathbf{M}$?

**Exercise 5.**
Repeat exercise 4, but now using two $D$ dimensional vectors $\mathbf{a} = (a_1, a_2, \ldots, a_D)^\top$ and $\mathbf{b} = (b_1, b_2, \ldots, b_D)^\top$.

What does this tell you about the rank of a matrix produced by the outer product of any vectors?

## 2 Partial Derivatives

**Exercise 6.**
Given the function

$$f(x, y, z) = 5x + 3xy - y^2 + 7xz + 2z^4 - yz\,,$$

evaluate the following expressions:

*a)* $\dfrac{\partial f}{\partial x}$

*b)* $\dfrac{\partial f}{\partial y}$

*c)* $\dfrac{\partial f}{\partial z}$

*d)* $\dfrac{\partial^2 f}{\partial z^2}$

*e)* $\dfrac{\partial^2 f}{\partial y \partial x}$

**Exercise 7.**
Using the *chain rule*

$$\frac{\partial f(g(x))}{\partial x} = f'(g(x))g'(x)$$

find the derivative of the following functions:

*a)* $f(x) = -2(x-4)^2$

*b)* $f(x) = (x+2)^5 - 3x$

*c)* $f(x) = (4x+1)^3 + 2$

*d)* $f(x) = (2x^2 - 4)^{-1}$

*e)* $f(x) = \dfrac{1}{5x+3}$

*f)* $f(x) = \sqrt{5x-2}$

*g)* $f(x) = 5\cos(x^3 - 2x)$

*h)* $f(x) = 3\ln(x^4 - x)$

### 2.1 Gradients

**Exercise 8.**
How is the gradient of a function $f$ (denoted by $\nabla f$) defined? Explain in words what the gradient means.

**Exercise 9.**
Find the gradient of the following functions. Find also the gradient and the magnitude of the gradient at the point $\mathbf{p} = (1, 2, 2)$.

*a*) $f(x, y, z) = x^2 + y^4 + z^3$

*b*) $g(x, y, z) = 3x^2 + y^2 - 2z^2 + 3xy - 2xz - 5yz$

*c*) $h(x, y, z) = 7x - y + 2z^3 - 3xz$

## 3 Completing the Square

**Exercise 10.**
Show that

$$f(x) = \exp\left[-\frac{1}{6}x^2 + \frac{1}{3}x - \frac{1}{6}\right]$$

is a Gaussian distribution by completing the square.

## 4 Likelihoods

**Exercise 11.**
Bayes' theorem can be written as

$$p(\mathbf{w} \,|\, \tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{x}} \,|\, \mathbf{w}) \times p(\mathbf{w})}{p(\tilde{\mathbf{x}})},$$

where $\tilde{\mathbf{x}} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_\ell\}$ is our data set and $\mathbf{w}$ is our parameter vector.

Label each term with its commonly known name (e.g., "likelihood" and "prior") and describe with your own words what they mean.

**Exercise 12.**
Explain why the likelihood of i.i.d. data points $\tilde{\mathbf{x}} = \{x_1, x_2, \ldots, x_\ell\}$ drawn from a distribution described by the parameters $\mathbf{w}$ can be written as

$$p(\tilde{\mathbf{x}} \,|\, \mathbf{w}) = \prod_{i=1}^{\ell} p(x_i \,|\, \mathbf{w}).$$

How would we normally write the expression for the likelihood in the case where the data point are drawn from a Gaussian distribution?

**Exercise 13.**
The log-likelihood is defined as $\ln(p(\tilde{\mathbf{x}} \,|\, \mathbf{w}))$. Show that the log-likelihood of a Gaussian can be written as

$$\ln p(\tilde{\mathbf{x}} \,|\, \mathbf{w}) = -\frac{\ell}{2}\ln\sigma^2 - \frac{\ell}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{i=1}^{\ell}(x_i - \mu)^2.$$

# 5 Gaussians

**Exercise 14.**
Show that the product of the two Gaussians

$$\mathcal{N}(\mathbf{x} \,|\, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_1|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1}(\mathbf{x}-\boldsymbol{\mu}_1)\right),$$
$$\mathcal{N}(\mathbf{x} \,|\, \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}_2|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_2)^\top \boldsymbol{\Sigma}_2^{-1}(\mathbf{x}-\boldsymbol{\mu}_2)\right)$$

is itself a Gaussian and find the mean $\boldsymbol{\mu}_3$ and covariance matrix $\boldsymbol{\Sigma}_3$.

# 6 Solutions

**Solution 1.**

*a)* $f(x) = \exp(ax + b)$

*b)* $f(x) = \exp(ax^2 + bx + c)$

*c)* $f(x, y) = \exp(ax^5 + by^{-3} + 7)$

*d)* $g(z) = \exp(8z - 3)$

*e)* $h(t) = \exp(-2t - 1)$

*f)* $h(t) = \exp(at - b)$

**Solution 2.**

*a)* $f(x) = \ln(abx)$

*b)* $f(x) = \ln(abx^2)$

*c)* $f(t) = \ln(a^{t+b})$

*d)* $g(x) = \ln(x/2)$

*e)* $g(y) = y$

*f)* $h(t) = 0$

**Solution 3.**

$$\|\mathbf{a}\| = \sqrt{1^2 + 4^2 + 5^2} = \sqrt{42} \approx 6.48$$

**Solution 4.**

$$\mathbf{a}\mathbf{b}^\mathsf{T} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \cdot [d, e, f] = \begin{bmatrix} ad & ae & af \\ bd & be & bf \\ cd & ce & cf \end{bmatrix}$$

Since the rows (or columns) are multiples of each other, i.e., there is only linear independent row (or column), the matrix has rank 1.

**Solution 5.**

$$\mathbf{a}\mathbf{b}^\mathsf{T} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_D \end{bmatrix} \cdot [b_1, b_2, \ldots, b_D] = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \cdots & a_1 b_D \\ a_2 b_1 & a_2 b_2 & \cdots & a_2 b_D \\ \vdots & \vdots & \ddots & \vdots \\ a_D b_1 & a_D b_2 & \cdots & a_D b_D \end{bmatrix}$$

Like in exercise 4, every row (or column) is a multiple of another row (or column), the matrix has rank 1. This means that every matrix, generated as the outer product of two arbitrary vectors, will always have rank 1. This means, among other things, that the matrix cannot be inverted.

**Solution 6.**

*a*) $\frac{\partial f}{\partial x} = 5 + 3y + 7z$

*b*) $\frac{\partial f}{\partial y} = 3x + 2y - z$

*c*) $\frac{\partial f}{\partial z} = 7x + 8z^3 - y$

*d*) $\frac{\partial^2 f}{\partial z^2} = 24z^2$

*e*) $\frac{\partial^2 f}{\partial y \partial x} = 3$

**Solution 7.**

*a*) $\frac{\partial}{\partial x} f = -4(x-4) = -4x + 16$

*b*) $\frac{\partial}{\partial x} f = 5(x+2)^4 - 3$

*c*) $\frac{\partial}{\partial x} f = 3(4x+1)^2 \cdot 4 = 12(4x+1)^2$

*d*) $\frac{\partial}{\partial x} f = -(2x^2+4)^{-2} \cdot 4x = -\frac{4x}{(2x^2+4)^2}$

*e*) $\frac{\partial}{\partial x} f = -(5x+3)^{-2} \cdot 5 = -\frac{5}{(5x+3)^2}$

*f*) $\frac{\partial}{\partial x} f = \frac{1}{2}(5x-2)^{-\frac{1}{2}} \cdot 5 = \frac{5}{2\sqrt{5x-2}}$

*g*) $\frac{\partial}{\partial x} f = -5\sin(x^3 - 2x)(3x^2 - 2)$

*h*) $\frac{\partial}{\partial x} f = \frac{3}{x^4 - x} \cdot (4x^3 - 1) = \frac{12x^3 - 3}{x^4 - x}$

**Solution 8.**

The gradient of a $D$ dimensional function $f$ is define as the vector $\nabla f = (\frac{\partial}{\partial x_1} f, \frac{\partial}{\partial x_2} f, \ldots, \frac{\partial}{\partial x_D} f)^\mathsf{T}$. For example, if $f$ a function in three dimensions ($x$, $y$ and $z$), $\nabla f = (\frac{\partial}{\partial x} f, \frac{\partial}{\partial y} f, \frac{\partial}{\partial z} f)^\mathsf{T}$.

As the gradient is the derivative of a function in each direction, it plays the same role as the slope of a curve in two dimensions. Each component of the gradient shows how much the function changes in the given direction. Intuitively, the gradient therefore forms a tangent to the function. The gradient points in the direction of the greatest change of the function, which makes it similar to the slope of two dimensional function, and the magnitude of the gradient is the slope in that direction.

**Solution 9.**

*a*) The gradient of $f(x, y, z)$ is

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} = \begin{bmatrix} 2x \\ 4y^3 \\ 3z^2 \end{bmatrix}$$

The gradient at $\mathbf{p}$ will be

$$\nabla f(\mathbf{p}) = \begin{bmatrix} 2 \cdot 1 \\ 4 \cdot 2^3 \\ 3 \cdot 2^2 \end{bmatrix} = \begin{bmatrix} 2 \\ 32 \\ 12 \end{bmatrix}$$

The magnitude of $\nabla f$ at $\mathbf{p}$ is

$$\|\nabla f(\mathbf{p})\| = \sqrt{2^2 + 32^2 + 12^2} = \sqrt{1172} \approx 34.23 \tag{1}$$

*b*) The gradient of $g(x, y, z)$ is

$$\nabla g = \begin{bmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \\ \frac{\partial g}{\partial z} \end{bmatrix} = \begin{bmatrix} 6x + 3y - 2z \\ 2y + 3x - 5z \\ -4z - 2x - 5y \end{bmatrix}$$

The gradient at $\mathbf{p}$ will be

$$\nabla f(\mathbf{p}) = \begin{bmatrix} 6 \cdot 1 + 3 \cdot 2 - 2 \cdot 2 \\ 2 \cdot 2 + 3 \cdot 1 - 5 \cdot 5 \\ -4 \cdot 2 - 2 \cdot 1 - 5 \cdot 2 \end{bmatrix} = \begin{bmatrix} 8 \\ -3 \\ -20 \end{bmatrix}$$

The magnitude of $\nabla g$ at $\mathbf{p}$ is

$$\|\nabla f(\mathbf{p})\| = \sqrt{8^2 + (-3)^2 + (-20)^2} = \sqrt{473} \approx 21.75 \tag{2}$$

*c*) The gradient of $h(x, y, z)$ is

$$\nabla h = \begin{bmatrix} \frac{\partial h}{\partial x} \\ \frac{\partial h}{\partial y} \\ \frac{\partial h}{\partial z} \end{bmatrix} = \begin{bmatrix} 7 - 3z \\ 1 \\ 6z^2 - 3x \end{bmatrix}$$

The gradient at $\mathbf{p}$ will be

$$\nabla f(\mathbf{p}) = \begin{bmatrix} 7 - 3 \cdot 2 \\ 1 \\ 6 \cdot 2^2 - 3 \cdot 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 21 \end{bmatrix}$$

The magnitude of $\nabla h$ at $\mathbf{p}$ is

$$\|\nabla f(\mathbf{p})\| = \sqrt{1^2 + 1^2 + 21^2} = \sqrt{443} \approx 21.05 \tag{3}$$

**Solution 10.**

Since we have an exponential of a quadratic term, we know that it can be written as a Gaussian distribution. Using completing the square (see the notes on Absalon), we can write the function

$$\begin{aligned} f(x) &= \exp\left[a\left(x + \frac{b}{2a}\right)^2 + a\left(\frac{c}{a} - \left(\frac{b}{2a}\right)^2\right)\right] && (4) \\ &= \exp\left[-\frac{1}{6}\left(x + \frac{\frac{1}{3}}{2 \cdot \left(-\frac{1}{6}\right)}\right) - \frac{1}{6}\left(\frac{\frac{1}{6}}{-\frac{1}{6}} - \left(\frac{\frac{1}{3}}{2 \cdot \left(-\frac{1}{6}\right)}\right)^2\right)\right] && (5) \\ &= \exp\left[-\frac{1}{6}(x-1)^2\right] && (6) \end{aligned}$$

**Solution 11.**
$p(\mathbf{w} \mid \tilde{\mathbf{x}})$ is the *posterior* and is the probability that a model with parameters ($\mathbf{w}$) would accurately describe the data $\tilde{\mathbf{x}}$.

$p(\tilde{\mathbf{x}} \mid \mathbf{w})$ is the *likelihood* and is the probability that a model with parameters $\mathbf{w}$ would produce the data $\tilde{\mathbf{x}}$ we observe. Note that this is not a probability distribution and the likelihood does not (necessarily) integrate to 1.

$p(\mathbf{w})$ is called the *prior* and is a subjective measure of how probable it is to get this particular set of parameters $\mathbf{w}$ out of all possible sets. Using the prior, we can restrict our parameters to, say, small values.

$p(\tilde{\mathbf{x}})$ is the *evidence* and is the probability of obtaining this particular set of data $\tilde{\mathbf{x}}$ from the underlying (true) distribution of data. This is impossible to determine, so the evidence is usually just defined as the normalisation constant that makes the posterior integrate to 1 (thereby making it a valid probability distribution).

**Solution 12.**
The joint probability of getting two random variables $x$ and $y$ (i.e., the probability of getting *both $x$ and $y$*) is

$$p(x, y) = p(y \mid x) p(x) .$$

If $x$ and $y$ are completely independent, this reduces to

$$p(x, y) = p(y) p(x) .$$

In the case of many independent variables $x_1, x_2, \ldots, x_\ell$, this becomes

$$p(\tilde{\mathbf{x}}) \equiv p(x_1, x_2, \ldots, x_\ell) = p(x_1) p(x_2) \cdots p(x_\ell) = \prod_{i=1}^{\ell} p(x_i) . \tag{7}$$

If the underlying distribution that produces the variables depends on some parameters $\mathbf{w}$, we would instead write

$$p(\tilde{\mathbf{x}} \mid \mathbf{w}) = \prod_{i=1}^{\ell} p(x_i \mid \mathbf{w}) . \tag{8}$$

In the particular case of an underlying Gaussian distribution, we would usually just write

$$p(\tilde{\mathbf{x}} \mid \mu, \sigma^2) = \prod_{i=1}^{\ell} \mathcal{N}(x_i \mid \mu, \sigma^2) .$$

**Solution 13.**
Remember that $\ln(ab) = \ln a + \ln b$, i.e., the logarithm of a product can be written as a sum of logarithms.

$$\begin{aligned}
\ln p(\tilde{\mathbf{x}} \mid \mathbf{w}) &= \ln \left[ \prod_{i=1}^{\ell} \mathcal{N}(x_i \mid \mu, \sigma^2) \right] \\
&= \sum_{i=1}^{\ell} \ln \left[ \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left( -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right) \right] \\
&= \sum_{i=1}^{\ell} \ln \left( \frac{1}{(2\pi\sigma^2)^{1/2}} \right) + \ln \left( \exp \left[ -\frac{1}{2\sigma^2} (x_i - \mu)^2 \right] \right) \\
&= \ell \ln \left( \frac{1}{(2\pi\sigma^2)^{1/2}} \right) \sum_{i=1}^{\ell} -\frac{1}{2\sigma^2} (x_i - \mu)^2 \\
&= -\frac{\ell}{2} \ln(2\pi) - \frac{\ell}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{\ell} (x_i - \mu)^2
\end{aligned}$$

**Solution 14.**
Solution to come. Sit tight!