Appendix A

A.1 Differentiation of real-valued functions of complex variables

Some properties of functions which map complex-valued variables to real-valued images is reviewed here. For details and applications of such an analysis, [57] is referred to. Suppose $\mathcal{A} \subset \mathbb{C}$ is an open set and a complex function $f(u) : \mathcal{A} \to \mathbb{C}$ is defined. The function $f(u)$ is said to be differentiable at $\hat{u} \in \mathcal{A}$ if its derivative at $\hat{u}$ defined as

$$\frac{\mathsf{d}}{\mathsf{d}u} f(u)\bigg|_{\hat{u}} = \lim_{u \to \hat{u}} \frac{f(u) - f(\hat{u})}{u - \hat{u}}, \tag{A.1}$$

exists. The function $f(u)$ is said to be analytical if the derivative exists for all $\hat{u} \in \mathcal{A}$. For analytical functions, the stationary points are located wherever

$$\frac{\mathsf{d}}{\mathsf{d}u} f(u) = 0. \tag{A.2}$$

The differential of an analytical $f(u)$ is given by

$$\mathsf{d}f(u) = \frac{\partial}{\partial u} f(u)\mathsf{d}u + \frac{\partial}{\partial \bar{u}} f(u)\mathsf{d}\bar{u}, \tag{A.3}$$

where $\bar{u} = u_1 - \mathsf{i}u_2$ is the complex conjugate of $u = u_1 + \mathsf{i}u_2$, where $u_1, u_2 \in \mathbb{R}$ and

$$\begin{aligned} \frac{\partial}{\partial u} &= \frac{1}{2}\left(\frac{\partial}{\partial u_1} - \mathsf{i}\frac{\partial}{\partial u_2}\right), \\ \frac{\partial}{\partial \bar{u}} &= \frac{1}{2}\left(\frac{\partial}{\partial u_1} + \mathsf{i}\frac{\partial}{\partial u_2}\right) \end{aligned} \tag{A.4}$$

are called Wirtinger derivatives. Also, note a direct consequence of (A.4) that

$$\frac{\partial}{\partial \bar{u}} u = \frac{\partial}{\partial u} \bar{u} = 0, \tag{A.5}$$

or $\bar{u}$ may be regarded as a constant when differentiating with respect to u, and vice-versa.

For any $f(u)$ that is not necessarily analytical, based on the condition (A.2), the stationary points may now be found by searching where

$$\mathsf{d}f(u) = 0. \tag{A.6}$$

Let $f(u) = f_1(u_1, u_2) + \mathrm{i} f_2(u_1, u_2)$, where $f_1, f_2 : \mathbb{R}^2 \to \mathbb{R}$. For $f(u)$ to be analytical, it is necessary that it satisfies the Cauchy-Riemann conditions

$$\frac{\partial}{\partial u_1} f_1 = \frac{\partial}{\partial u_2} f_2, \qquad \frac{\partial}{\partial u_2} f_1 = -\frac{\partial}{\partial u_1} f_2. \tag{A.7}$$

Now, focus the situation in which $f(u) : \mathcal{A} \to \mathbb{R}$. Firstly, the conditions (A.7) show that $f(u)$ is analytical if and only if $f(u)$ is constant. Secondly, $\mathrm{d}f = 2\Re(\frac{\partial}{\partial u} f(u)\mathrm{d}u) = 2\Re(\frac{\partial}{\partial \bar{u}} f(u)\mathrm{d}\bar{u})$, which vanishes if and only if

$$\frac{\partial}{\partial u} f(u) = 0. \tag{A.8}$$

Hence, for finding the stationary points of a non-analytical function, the trick involves writing the differential in the form of (A.3) and set the term corresponding to $\frac{\partial}{\partial u} f(u)$ to zero.

In the multivariate case [71, 59], for the complex-valued function $f(u) : \mathcal{A} \subset \mathbb{C}$ with $\mathcal{A} \subset \mathbb{C}^r$,

$$\mathrm{d}f = \frac{\partial}{\partial u'} f(u)\,\mathrm{d}u + \frac{\partial}{\partial u^*} f(u)\,\mathrm{d}(\bar{u}), \tag{A.9}$$

where $u^* \equiv \bar{u}'$ is the conjugate transpose of u. It then easily follows that the differential $\mathrm{d}f$ of a real-valued function $f(u) : \mathcal{A} \to \mathbb{R}\, \forall u \in \mathcal{A} \subset \mathbb{C}^n$ vanishes if and only if the Wirtinger derivative is zero, i.e.,

$$\mathrm{d}f(u) = 0 \Leftrightarrow \frac{\partial}{\partial u} f(u) = 0. \tag{A.10}$$

Appendix B

B.1 Certain details of the EM Algorithm

To enable a smooth reading of the EM Algorithm developed in Section 3.5, certain details are let to reside separately. They are elucidated here:

B.1.1 Log-likelihood as summation of logarithms

The following lemma is well-known; refer §16.5.4 of [30]:

> **Lemma B.1.** *Suppose that* $u_1, \ldots, u_m$ *are points in the interval* $\mathcal{U}$ *and* $c_1, \ldots, c_m \geq 0$ *are such that* $\sum_{l=1}^{m} c_l = 1$ *and* f *is a concave function in* $\mathcal{U}$. *According to* ***Jensen's inequality*** $f(c_1 u_1 + \cdots + c_m u_m) \geq c_1 f(u_1) + \cdots + c_m f(u_m)$.

With $f \leftarrow \log_e$, $c_l \leftarrow g(x_l)$, and $u_l \leftarrow \mathsf{p}^{\mathsf{y},\mathsf{x}_l|\theta}(\mathcal{D}, x_l \mid \theta)/g(x_l)$, (3.32) is got.

B.1.2 Decomposition of the complete log-likelihood

Using Theorem B.1, $\mathsf{p}^{\mathsf{y},\mathsf{x}|\theta}(\mathcal{D}, x \mid \theta) = \mathsf{p}^{\mathsf{y}|\theta}(\mathcal{D} \mid \theta)\, \mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid \mathcal{D}, \theta)$ is obtained. Hence, the right side of (3.32) may be factorized so that

$$\mathsf{L}(\theta, g) \geq \sum_{\mathsf{x}} g(x) \log_e \mathsf{p}^{\mathsf{y}|\theta}(\mathcal{D} \mid \theta) + \sum_{\mathsf{x}} g(x) \log_e \frac{\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid \mathcal{D}, \theta)}{g(x)},$$

where the first term reduces to $\mathsf{L}(\theta)$ due to (3.5) and (3.30).

B.1.3 Maximization of an expectation

If $\hat{g}_i$ of (3.35) is substituted in (3.32)

$$\mathsf{L}(\theta, \hat{g}_i) = \sum_{x} \mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid \mathcal{D}, \theta_i) \log_e \frac{\mathsf{p}^{\mathsf{y},\mathsf{x}|\theta}(\mathcal{D}, x \mid \theta)}{\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid \mathcal{D}, \theta_i)},$$

where the denominator in the logarithm being independent of θ may be eliminated. As a result, (3.37) boils down to

$$\begin{aligned}\hat{\theta}_{i+1} &= \underset{\theta_i}{\operatorname{argmax}}\ \mathsf{L}(\theta_i, \hat{g}_i)\\ &= \underset{\theta_i}{\operatorname{argmax}} \sum_x \mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid \mathcal{D}, \theta_i) \log_e \mathsf{p}^{\mathsf{y},\mathsf{x}|\theta}(\mathcal{D}, x \mid \theta_i),\\ &= \underset{\theta_i}{\operatorname{argmax}}\ \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}\big[\log_e \mathsf{p}^{\mathsf{y},\mathsf{x}|\theta}(\mathcal{D}, x \mid \theta_i)\big].\end{aligned}$$

B.2 Posterior density with a Gaussian prior

Refer [113, 74] and §6.2 of [95] for the following theorem:

Theorem B.1. *According to the **Bayes theorem** for continuous probability density functions, the conditional distribution of a random variable* y *with any realization* y *given a set of random variables* x *with any realization* x *is related to the conditional distribution of* x *given* y *according to*

$$\mathsf{p}^{\mathsf{x}}(x)\mathsf{p}^{\mathsf{y}|\mathsf{x}}(y \mid x) = \mathsf{p}^{\mathsf{y}}(y)\, \mathsf{p}^{\mathsf{x}|\mathsf{y}}(x \mid y) \equiv \mathsf{p}^{\mathsf{y},\mathsf{x}}(y, x).$$

Due to Theorem B.1, $\mathsf{p}^{\mathsf{x}|\mathsf{y}}(x \mid y) = \frac{\mathsf{p}^{\mathsf{x}}(x)\mathsf{p}^{\mathsf{y}|\mathsf{x}}(y|x)}{\mathsf{p}^{\mathsf{y}}(y)}$; so, given the parameters θ, it follows that $\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid y, \theta) = \frac{\mathsf{p}^{\mathsf{x}|\theta}(x|\theta)\mathsf{p}^{\mathsf{y}|\mathsf{x}}(y|x,\theta)}{\mathsf{p}^{\mathsf{y}}(y|\theta)}$. While a Gaussian has been accepted for the denominator $\mathsf{p}^{\mathsf{y}}(y \mid \theta)$ according to (3.18), $\mathsf{p}^{\mathsf{y}|\mathsf{x}}(y \mid x, \theta)$ in the numerator is also a Gaussian as per (3.39). Assuming yet another Gaussian for

$$\mathsf{p}^{\mathsf{x}|\theta}(x \mid \theta) = \mathsf{p}^{\mathsf{x}}(x) = \mathcal{N}(x \mid 0, I_q).$$

Therefore,

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid y, \theta) = \frac{\mathcal{N}(x \mid 0, I_q)\mathcal{N}(y \mid Wx, \Gamma^{\mathsf{z}})}{\mathcal{N}(y \mid \mu^{\mathsf{y}}, \Gamma^{\mathsf{y}})}. \tag{B.1}$$

Suppose $c_1, \ldots, c_4$ are factors independent of x such that

$$\begin{aligned}\mathcal{N}(x \mid 0, I_q) &= c_1 \exp(-0.5\, x'x),\\ \mathcal{N}(y \mid Wx, \Gamma^{\mathsf{z}}) &= c_2 \exp(-0.5\, x'W'(\Gamma^{\mathsf{z}})^{-1}Wx + x'W'(\Gamma^{\mathsf{z}})^{-1}y),\\ \mathcal{N}(y \mid \mu^{\mathsf{y}}, \Gamma^{\mathsf{y}}) &= c_3,\\ c_4 &= \frac{c_1 c_2}{c_3}.\end{aligned}$$

Then, (B.1) may be written as

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid y, \theta) = c_4 \exp(-0.5\, x'\Omega^{-1}x + x'\Omega^{-1}\Omega W'(\Gamma^{\mathsf{z}})^{-1}y),$$

where

$$\Omega^{-1} = I_q + W'(\Gamma^{\mathsf{z}})^{-1}W.$$

The probability density function of a Gaussian ξ with mean a and covariance matrix B may be written as $\mathcal{N}(\xi \mid a, B) = c\,\mathsf{exp}(-0.5\,\xi' B^{-1}\xi + \xi' B^{-1} a)$, where c is a factor independent of ξ. Thus, $\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid y, \theta)$ is a Gaussian with mean $\Omega W'(\Gamma^{\mathsf{z}})^{-1} y$ and covariance matrix Ω. It can be seen that

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(x \mid y, \theta) = \mathcal{N}(x \mid \Omega W'(\Gamma^{\mathsf{z}})^{-1} y, \Omega).$$

B.3 Posterior density with a complex Gaussian prior

The extension of Section B.2 to complex Gaussian densities is straightforward. In that order of equations and interpretations therein, the following relations hold:

$$\mathsf{p}^{\mathsf{x}|\theta}(\mathbf{x} \mid \theta) = \mathcal{N}_{\mathbb{C}}(\mathbf{x} \mid 0, I_q)$$

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(\mathbf{x} \mid \mathbf{y}, \theta) = \frac{\mathcal{N}_{\mathbb{C}}(\mathbf{x} \mid 0, I_q)\mathcal{N}_{\mathbb{C}}(\mathbf{y} \mid \mathbf{W}\mathbf{x}, \mathcal{S}^{\mathsf{z}})}{\mathcal{N}_{\mathbb{C}}(\mathbf{y} \mid 0, \mathcal{S}^{\mathsf{y}})}. \tag{B.2}$$

Suppose $c_1, \ldots, c_4$ are factors independent of $\mathbf{x}$ such that

$$\mathcal{N}_{\mathbb{C}}(\mathbf{x} \mid 0, I_q) = c_1\,\mathsf{exp}(-\mathbf{x}^*\mathbf{x}),$$

$$\mathcal{N}_{\mathbb{C}}(\mathbf{y} \mid \mathbf{W}\mathbf{x}, \mathcal{S}^{\mathsf{z}}) = c_2\,\mathsf{exp}(-\mathbf{x}^*\mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{W}\mathbf{x} + 2\Re(\mathbf{x}^*\mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{y})),$$

$$\mathcal{N}_{\mathbb{C}}(\mathbf{y} \mid 0, \mathcal{S}^{\mathsf{y}}) = c_3,$$

$$c_4 = \frac{c_1 c_2}{c_3}.$$

Then, (B.2) may be written using

$$\mathbf{\Omega}^{-1} = I_q + \mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{W}.$$

as

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(\mathbf{x} \mid \mathbf{y}, \theta) = c_4\,\mathsf{exp}(-\mathbf{x}^*\mathbf{\Omega}^{-1}\mathbf{x} + 2\,\Re(\mathbf{x}^*\mathbf{\Omega}^{-1}\mathbf{\Omega}\mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{y}))$$

The probability density function of a complex Gaussian ξ with mean a and covariance matrix B may be written as $\mathcal{N}_{\mathbb{C}}(\xi \mid a, B) = c\,\mathsf{exp}(-\xi^* B^{-1}\xi + 2\Re(\xi^* B^{-1} a))$, where c consists of the normalization factor of the distribution independent of ξ. This shows that $\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(\mathbf{x} \mid \mathbf{y}, \theta)$ above is a complex Gaussian with mean $\mathbf{\Omega}\mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{y}$ and covariance matrix $\mathbf{\Omega}$, i.e.,

$$\mathsf{p}^{\mathsf{x}|\mathsf{y},\theta}(\mathbf{x} \mid \mathbf{y}, \theta) = \mathcal{N}_{\mathbb{C}}(\mathbf{x} \mid \mathbf{\Omega}\mathbf{W}^*(\mathcal{S}^{\mathsf{z}})^{-1}\mathbf{y}, \mathbf{\Omega}).$$