Chapter 5

Maximum likelihood commonalities

The objective of this chapter is to solve the maximization problem defined as part of the spectral factor model in Section 4.2. That problem refers to maximizing the commonalities retained by the latent spectral factor transformation. It will be shown that its solution requires estimating the **maximum likelihood spectral density function**. Two methods are developed to arrive at the maximum likelihood spectral density function estimates: The first method is analytical and is the topic of Section 5.1; it gains traction from the estimation procedures summarized in Section 3.4. The second method discussed in Section 5.2 is iterative; it is along the lines of the EM algorithm presented in Section 3.6.

In Section 5.1, as part of the analytical method, optimal parameters of the spectral factor model are made available in (5.10) and (5.13). In order to arrive at those results, the expression for the log-likelihood function of the spectral factor model is written. Due to difficulties in maximizing such a real-valued function of complex-valued parameters, Wirtinger relaxation rules of complex differentiation are sought. Such an approach gives the relation (5.5) connecting the spectral density functions of the latent and the idiosyncratic processes to the sample measured spectral density function. Sadly, it evades delivering a unique solution. Therefore and subsequently, a much restricted class of maximum likelihood solutions is pursued where the commonalities will be maximized as well. Towards the end of that pursuit, the low-rank approximation technique of Section 3.4.2 is used to arrive at a suitable solution.

In Section 5.2, the objective is to iteratively solve the commonality maximization problem defined as part of the spectral factor model in Section 4.2. The optimal parameters of the spectral factor model are made available in (5.33) and (5.34). Just as with the analytical method in Section 5.1, first the maximum likelihood parameters of the spectral factor model are obtained; here it is done iteratively using the EM algorithm. In doing so, the line of the estimation approach in Section 3.6 for Section 5.2 is towed by which the definition of the 'E' and 'M' steps are laid out. For this purpose, the formulae for the *a posteriori* expectation and the maximum likelihood parameters are carried out just as they were derived in Sections 3.7.1 and 3.7.2. However, the analysis is tedious because of the non-analytic nature of the real-valued log-likelihood function of complex-valued parameters. As in Section 5.1, this difficulty is overcame by employing Wirtinger relaxations. The equations (5.25) and (5.29) give the maximum likelihood parameters of the spectral factor model at each iteration of the EM algorithm. Once the EM algorithm has converged, the parameters that maximize the commonalities are found in Section 5.2.3 using the idea of an efficient unbiased estimator reviewed in Section 3.2.

Note 5.1. *For the analysis in this chapter, the focus is on any one and only one target frequency in the set of target frequencies obtained on application of Theorem 2.5. Hence, for brevity of notations in this chapter, the index specifying different subbands will be dropped. Therefore, the sans-serif script without any subscripts as in* y *will be used to refer to the discrete Fourier transform vector random variable at the target frequency of interest. As a result, the spectral density function at the target frequency is simply* S^{y} *and the transformation matrix is* $\mathbf{W}$.

Maximum likelihood estimation of linear processes in time-domain

The attempt in this thesis is to use the principle of maximum likelihood to estimate parameters of the dynamic factor model in the frequency domain. Despite the challenges posed by complex-valued parameters of the model, such an approach was motivated by the established route of maximum likelihood in factor analysis.

An alternative route that should easily be motivated by the maximum likelihood principle is the estimation of the dynamic factor model in the time domain [78]; however, it is not pursued in this thesis. It involves expressing the large sample approximation of the likelihood function in terms of finite-order vector autoregressive moving average process parameters; the maximum likelihood parameters are known to be consistent and asymptotically Gaussian. The derivative of the likelihood function with respect to the parameters are typically non-linear. Hence, iterative algorithms such as Newton - Raphson scoring algorithm [8] or state-space Expectation - Maximization algorithm [83] are used to maximize the log-likelihood. These iterative procedures in the time-domain for vector autoregressive moving average processes are complicated owing to a large set of parameters requiring reliable initial values as well as convergence issues requiring robust estimates of model orders; refer Chapter 12 of [78]. The efficacy of adopting these methods to dynamic factor model estimation in the time-domain is yet to be seen.

On the other hand, the frequency-domain method as presented in this thesis exploits proven methodologies to solve the estimation problem. The analytical approach of Section 5.1 offers an intuitive computationally stable closed-form solution; it uses low-rank approximation theorem and Weyl's theorem to arrive at maximum commonalities parameters. The iterative approach of Section 5.2 uses the EM algorithm for complex-Gaussian estimation and Gauss-Markov theorem. Beyond the known-issue of local minima, it does not suffer from over-parameterization and, as presented in this thesis, is computational stable for Gaussian factor model estimation [14].

The following situates the developments in this chapter with respect to the state-of-the-art:

- ▷ The analytical solution for spectral factor model is derived in (5.10) using low-rank approximation theorem.
  The solution, which involves the principal components of the sample spectral

density function, coincides with that of the the projection theorem solution of [36]; they have a motivation and approach to dynamic factor model not dependent on commonalities.

- ▷ An iterative solution for spectral factor model is derived in Section 5.2 using the Expectation - Maximization algorithm. The converged maximum likelihood parameters in Section 5.2.3 that maximally inherit the commonalities are extracted by applying the Gauss - Markov theorem.
  Iterative solutions recommended by [104] and [100] were based on Fletcher-Powell-Davidon numerical methods.

- ▷ Mild cross-correlation property of the difference between the maximally inherited commonalities and the measured variables in Property 5.1 is obtained via Weyl's theorem.
  In [37], a similar result is obtained via "monotone convergence theorem".

- ▷ Wirtinger relaxations are used for maximizing log-likelihood.
  Relations (5.4) - (5.6) in Section 5.1 states a well-known fact that the sample spectral density maximizes the log-likelihood; e.g., [104] calls it the "unobservable index model". They are retold here using Wirtinger relaxations to emphasize nonexistence, in the Cauchy-Riemann sense, of a non-trivial derivative of the real-valued log-likelihood function of complex-valued variables. The relaxations are introduced in the very familiar setting of Section 5.1 in anticipation of its use in Section 5.2. An alternative of using the isomorphic relations of a complex-Gaussian with that of a real-Gaussian as discussed in Section 2.5 could prove tedious for the purposes in Section 5.2.

5.1 Analytical estimation of maximum likelihood commonalities

Note 5.2. *Since the selected target frequency represents a subband of frequencies near it, the realization of* y *corresponding to the l-th frequency sample within the subband near the target frequency is referred to by $\mathbf{y}(\omega_l)$.*

In Theorem 2.5, an asymptotic property of the discrete Fourier transform was reviewed. It involved treating the discrete Fourier transform at a target frequency ω as a complex vector random variable y whose realizations are asymptotically the discrete Fourier transform samples $\mathbf{y}(\omega_l) \in \mathbb{C}^r$ at appropriately spaced frequencies ω_l near ω. It was observed there that these samples may be thought to have been generated from a complex Gaussian probability density

$$\mathsf{p}^{\mathsf{y}}(\mathbf{y}(\omega_l)) = \pi^{-r}(\det(\mathsf{S}^{\mathsf{y}}))^{-1} \exp(-\mathbf{y}'(\omega_l)(\mathsf{S}^{\mathsf{y}})^{-1}\mathbf{y}(\omega_l)), \tag{5.1}$$

where $\mathsf{S}^{\mathsf{y}} \in \mathbb{C}^{r\times r}$ is the spectral density function at frequency ω. For n such discrete Fourier transform samples $\mathbf{y}(\omega_l)$, $l = 1,\ldots,n$, their log-likelihood function may be written as $-rn\log(\pi) - n\log(\det(\mathsf{S}^{\mathsf{y}})) - \sum_{l=1}^{n} \mathbf{y}'(\omega_l)(\mathsf{S}^{\mathsf{y}})^{-1}\mathbf{y}(\omega_l)$. The terms which are independent of S^{y} may be discarded and the effective log-likelihood is written as

$$\mathsf{L}(\mathsf{S}^{\mathsf{y}}) = -\log(\det(\mathsf{S}^{\mathsf{y}})) - \mathrm{tr}((\mathsf{S}^{\mathsf{y}})^{-1}\check{\mathsf{S}}^{\mathsf{y}}), \tag{5.2}$$

where $\check{\mathsf{S}}^{\mathsf{y}} \in \mathbb{C}^{r\times r}$ is the sample spectral density function as per (2.30). Note that the inner product of two vectors is converted to the trace of their outer product.

The log-likelihood function $\mathsf{L}(\mathsf{S}^{\mathsf{y}})$ is a real-valued function of complex valued variables in S^{y}. Hence, it is a non-analytical function and its stationary points have to be found from its vanishing differential $\mathsf{dL}(\mathsf{S}^{\mathsf{y}})$. Presenting the details of deriving the differentials of common real-valued functions of complex-valued matrices is skipped. A comprehensive treatment starting from the basic idea mentioned in Appendix A.1 to a full-fledged multivariate complex calculus is beyond the scope of this thesis. Instead, among many good references, the reader is referred to [57]. Referring to Tables II and V of [57], it is easy to verify that $\mathsf{d}\log(\det(\mathsf{S}^{\mathsf{y}})) = \mathsf{tr}((\mathsf{S}^{\mathsf{y}})^{-1}\mathsf{dS}^{\mathsf{y}})$ and $\mathsf{d}\,\mathsf{tr}((\mathsf{S}^{\mathsf{y}})^{-1}\check{\mathsf{S}}^{\mathsf{y}}) = -\mathsf{tr}((\mathsf{S}^{\mathsf{y}})^{-1}\check{\mathsf{S}}^{\mathsf{y}}(\mathsf{S}^{\mathsf{y}})^{-1}\mathsf{dS}^{\mathsf{y}})$, and their sum may be written as

$$\mathsf{dL}(\mathsf{S}^{\mathsf{y}}) = -\mathsf{tr}(((\mathsf{S}^{\mathsf{y}})^{-1} - (\mathsf{S}^{\mathsf{y}})^{-1}\check{\mathsf{S}}^{\mathsf{y}}(\mathsf{S}^{\mathsf{y}})^{-1})\mathsf{dS}^{\mathsf{y}}).$$

Based on this differential and from the trace form of the differentials in Table III of [57],

$$\frac{\partial}{\partial \mathsf{S}^{\mathsf{y}}}\mathsf{L}(\mathsf{S}^{\mathsf{y}}) = -(\mathsf{S}^{\mathsf{y}})^{-1} + (\mathsf{S}^{\mathsf{y}})^{-1}\check{\mathsf{S}}^{\mathsf{y}}(\mathsf{S}^{\mathsf{y}})^{-1}. \tag{5.3}$$

As mentioned in Appendix A.1, the stationary points of $\mathsf{L}(\mathsf{S}^{\mathsf{y}})$ occur wherever $\mathsf{dL}(\mathsf{S}^{\mathsf{y}})$ vanishes. Since $(\mathsf{S}^{\mathsf{y}})^{-1} = 0$ is prohibited for the existence of S^{y}, the maximum likelihood solution is

$$\widehat{\mathsf{S}}^{\mathsf{y}} = \check{\mathsf{S}}^{\mathsf{y}} \tag{5.4}$$

wherever $\frac{\partial}{\partial \mathsf{S}^{\mathsf{y}}}\mathsf{L}(\mathsf{S}^{\mathsf{y}}) = 0$. Now substitute (4.12) in the maximum likelihood solution for S^{y} in (5.4); it follows that

$$\mathsf{S}^{\mathsf{v}} + \mathsf{S}^{\mathsf{z}} = \check{\mathsf{S}}^{\mathsf{y}}, \tag{5.5}$$

where the check denotes the sample estimate of the spectral density function. Based on (4.11) the maximum likelihood estimates may be further rewritten as

$$\mathbf{W}\mathbf{W}^* + \mathsf{S}^{\mathsf{z}} = \check{\mathsf{S}}^{\mathsf{y}}. \tag{5.6}$$

Since maximum likelihood solution for the parameters $\mathbf{W}$ and S^{z} have to be gleaned from just one relation in (5.6), there will not be any unique solution.

In order to find the parameters that maximize the commonalities of y amongst the maximum likelihood parameters $\mathbf{W}$ and S^{z} of (5.6), further restrictions on the quality of the solutions will have to be imposed. Recall that Definition 4.1 of the commonalities led the formulation of (4.19) which meant v will inherit the covariation in y maximally according to relation (4.18).

However, note that the trivial solution that the diagonal matrix $\mathsf{S}^{\mathsf{z}}(\omega) = 0\ \forall\omega \in [-\pi,\pi]$ and $\check{\mathsf{S}}^{\mathsf{y}}(\omega) = \mathsf{S}^{\mathsf{v}}(\omega)$ is forbidden because $\mathsf{rank}(\check{\mathsf{S}}^{\mathsf{y}}) = r \neq \mathsf{rank}(\mathsf{S}^{\mathsf{v}}) = q$.

Parameters due to commonalities

Note that the function to be minimized in (4.19) is nonnegative for every ω in the integral in (4.18). Hence, $\|\check{\mathsf{S}}^{\mathsf{y}}(\omega) - \mathsf{S}^{\mathsf{v}}(\omega)\|_F^2$ may be minimized for each ω individually and specifying the variable ω may be dropped for brevity. Therefore, the maximum commonalities maximum likelihood solution should solve

$$\widetilde{\mathbf{W}} = \underset{\mathbf{W}}{\text{argmin}}\ \|\check{\mathsf{S}}^{\mathsf{y}} - \mathsf{S}^{\mathsf{v}}\|_F^2, \tag{5.7}$$
$$\text{rank}(\mathsf{S}^{\mathsf{v}}) = q < \text{rank}(\check{\mathsf{S}}^{\mathsf{y}}) = r.$$

Recall that according to Theorem 3.1, for the eigenvalue decomposition

$$\check{\mathsf{S}}^{\mathsf{y}} = U\text{diag}(\lambda_1, \cdots, \lambda_r)U^*, \tag{5.8}$$

where $U = [u_1 \cdots u_r]$ is unitary and $\lambda_1 \geq \lambda_r > 0$ are the eigenvalues of $\check{\mathsf{S}}^{\mathsf{y}}$, the best q rank approximation in the Frobenius norm sense is

$$\widetilde{\mathsf{S}}^{\mathsf{v}} = [u_1 \cdots u_q]\,\text{diag}(\lambda_1, \cdots, \lambda_q)\,[u_1 \cdots u_q]^*. \tag{5.9}$$

Then it is straightforward to observe that for $\mathsf{S}^{\mathsf{v}} = \mathbf{W}\mathbf{W}^*$ in (4.11), a possible decomposition for the optimal $\mathbf{W}$ is

$$\widetilde{\mathbf{W}} = [u_1 \cdots u_q]\,\text{diag}(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_q}). \tag{5.10}$$

Comparing the result (5.10) with that of classical principal components analysis [88], it can be seen that columns of $\widetilde{\mathbf{W}}$ are seeking directions in which the sample measured variances are maximally retained.

Properties of non-commonalities

Note at this juncture from (5.5) that

$$\text{tr}(\widetilde{\mathsf{S}}^{\mathsf{v}}) = \sum_{i=1}^{q} \lambda_i, \tag{5.11}$$

which enables

$$\text{tr}(\widetilde{\mathsf{S}}^{\mathsf{z}}) = \text{tr}(\check{\mathsf{S}}^{\mathsf{y}}) - \text{tr}(\widetilde{\mathsf{S}}^{\mathsf{v}}) = \sum_{i=q+1}^{r} \lambda_i. \tag{5.12}$$

For the following lemma, refer Chapter 1 §4.4 of [116]:

Theorem 5.1. *(Weyl's theorem) For $A, B, C \in \mathbb{C}^{r\times r}$ whose eigenvalues are $a_1 \geq \cdots \geq a_r$, $b_1 \geq \cdots \geq b_r$, and $c_1 \geq \cdots \geq c_r$, respectively, if $A = B + C$, then $b_i + c_1 \geq a_i \geq b_i + c_r$.*

Let $\check{\mathsf{S}}^{\mathsf{y}}$, $\widetilde{\mathsf{S}}^{\mathsf{v}}$, $\widetilde{\mathsf{S}}^{\mathsf{z}}$ correspond to A, B, C, respectively, in Theorem 5.1. Recall that the least $r - q$ eigenvalues of $\widetilde{\mathsf{S}}^{\mathsf{v}}$ are equal to zero. Then, it follows that any $\widetilde{\mathsf{S}}^{\mathsf{z}}$ satisfying (5.12) may be chosen such that

$$\text{tr}(\widetilde{\mathsf{S}}^{\mathsf{v}}) > c_1 \geq (r - q)^{-1}\text{tr}(\widetilde{\mathsf{S}}^{\mathsf{z}}) \geq c_r > 0. \tag{5.13}$$

This inequality establishes the following property for z while preserving the Frobenius norm criterion for the inheritance of commonalities defined through 5.7.

Property 5.1. *The variables forming the non-commonalities* $\{z_t\}$ *of the dynamic factor model in (4.2) may be 'mildly cross-correlated' as per (5.12) and (5.13).*

Property 5.1 suggests that z need not be strictly idiosyncratic.

Suppose the discrete Fourier transform components $\mathcal{D} = \{\mathbf{y}(\omega_l)\}$, $l = 1, \ldots n$ within a subband as per Algorithm 1 are obtained. The solution proposed in Algorithm 3 provides the analytical solution of the spectral factor model within a subband.

Algorithm 3: Analytical solution for the spectral factor model in a subband

Input: $\mathcal{D} = \{\mathbf{y}(\omega_l)\}, l = 1, \ldots n$
Output: $\widetilde{\mathbf{W}}$
estimate $\widehat{\mathsf{S}}^{\mathsf{y}}$ using (5.4) and (2.30);
compute pairs (λ_k, u_k) $\forall k = 1, \ldots, r$ using (5.8);
estimate $\widetilde{\mathbf{W}}$ as in (5.10);

5.2 Iterative estimation of maximum likelihood commonalities

In Section 3.6, an iterative solution for the parameters of the classical factor model of (3.11) was developed. Based on the EM algorithm presented therein, in this section, an iterative procedure for the estimation of the maximum likelihood parameters which is also enforced to maximally inherit the measured commonalities will be developed. Such a motivation to do so is due to the similarity of the relations of the classical factor model in (3.11) and the spectral factor model (4.15). This similarity is obvious if it is supposed that the parameters of the factor model are $\theta \triangleq \{\mathbf{W}, \mathsf{S}^{\mathsf{z}}\}$ and that the random vectors y and z at a target frequency realize $\mathbf{y}$ and $\mathbf{z}$ at nearby frequencies according to Theorem 2.5.

Note 5.3. *As in the previous section, the realization of* y *corresponding to the* l*-th frequency sample within the subband near the target frequency is denoted by* $\mathbf{y}(\omega_l)$*. In addition, in this section,* $\mathsf{S}^{\mathsf{y}}_i$ *and* $\mathbf{W}_i$ *are used to refer to the* i*-th iterative estimate of the spectral density function* S^{y} *and the transformation matrix* $\mathbf{W}$ *at the target frequency, respectively.*

As in Section 3.6, first notice that the spectral factor model equivalent of (3.39) is

$$\mathsf{p}^{\mathsf{y}|\mathsf{x},\theta}(\mathbf{y} \mid \mathsf{x}, \theta) = \mathcal{N}_{\mathbb{C}}(\mathbf{y} \mid \mathbf{W}\mathsf{x}, \mathsf{S}^{\mathsf{z}}). \tag{5.14}$$

Let a dataset $\mathcal{D}$ render the discrete Fourier transform samples $\mathbf{y}(\omega_l)$, $l = 1, \ldots, n$ at frequencies within the subband represented by the random vector y. At the target frequency under consideration, the likelihood of $\mathcal{D}$ to correspond to the spectral factor model is

$$\mathsf{p}^{\mathsf{y}|\mathsf{x},\theta}(\mathcal{D} \mid \mathsf{x}, \theta) = \prod_{l=1}^{n} \mathsf{p}^{\mathsf{y}|\mathsf{x},\theta}(\mathbf{y}(\omega_l) \mid \mathsf{x}, \theta). \tag{5.15}$$

Now, in line with (3.41), (3.43), and (3.42), the direct extension of the estimation of the spectral factor model parameters in the i-th iteration of the EM algorithm gives:

$$\hat{\theta}_{i+1} = \underset{\theta_i}{\mathsf{argmax}}\ \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}\big[\log_e \mathsf{p}^{\mathsf{y}|\mathsf{x},\theta}(\mathcal{D} \mid \mathsf{x}, \theta_i)\big] \tag{5.16}$$

and for

$$\begin{aligned} \theta_i &\triangleq \{\mathbf{W}_i, \mathsf{S}^{\mathsf{z}}_i\}, \\ \hat{\theta}_{i+1} &= \underset{\theta_i}{\mathsf{argmax}}\ \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})], \\ \mathsf{f}(\theta_i, \mathsf{x}) &\triangleq \sum_{l=1}^{n} \log_e \mathcal{N}_{\mathbb{C}}(\mathbf{y}(\omega_l) \mid \mathbf{W}_i \mathsf{x}, \mathsf{S}^{\mathsf{z}}_i). \end{aligned}$$

Expanding $\mathsf{f}(\theta_i, \mathsf{x})$ will lead to terms in x and $\mathsf{x}\mathsf{x}^*$. So, as with (3.46), first define

$$\begin{aligned} \langle \mathbf{x} \rangle_{i,l} &\triangleq \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{x} \mid \mathbf{y}(\omega_l), \theta_i], \\ \langle \mathbf{x}\mathbf{x}^* \rangle_{i,l} &\triangleq \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{x}\mathsf{x}^* \mid \mathbf{y}(\omega_l), \theta_i]. \end{aligned} \tag{5.17}$$

Note that $\langle \mathbf{x} \rangle_{i,l} \in \mathbb{C}^q$ and $\langle \mathbf{x}\mathbf{x}^* \rangle_{i,l} \in \mathbb{C}^{q \times q}$; whereas their estimation will define the E-step of the EM-algorithm. Then, as in (3.51), it may be written that

$$\mathbf{W}_{i+1} = \underset{\mathbf{W}_i}{\mathsf{arg}} \left(\frac{\partial}{\partial \mathbf{W}_i} \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})] = 0 \right). \tag{5.18}$$

Similarly, as in (3.53),

$$\mathsf{S}^{\mathsf{z}}_{i+1} = \underset{\mathsf{S}^{\mathsf{z}}_i}{\mathsf{arg}} \left(\frac{\partial}{\partial \mathsf{S}^{\mathsf{z}}_i} \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})] = 0 \right). \tag{5.19}$$

These optimizations complete the M-step of the EM-algorithm.

Hence, starting from initial guesses, the i-th iteration alternates between:

1. **Expectation-step** Evaluate $\langle \mathbf{x} \rangle_{i,l}$ and $\langle \mathbf{x}\mathbf{x}^* \rangle_{i,l}$ using (5.17), and
2. **Maximization-step** Update $\mathbf{W}_{i+1}$ using (5.18) and $\mathsf{S}^{\mathsf{z}}_{i+1}$ using (5.19).

It is clear that the EM algorithm leads to non-unique maximum likelihood solutions depending on the starting conditions.

5.2.1 EM steps and form of the maximum likelihood parameters

As in the previous sections, note that ω_l corresponds to the l-th frequency sample within the subband near the target frequency. Also, $\mathsf{S}_i^{\mathsf{y}}$ and $\mathbf{W}_i$ refer to the i-th iterative estimate of the spectral density function S^{y} and the transformation matrix $\mathbf{W}$ at the target frequency, respectively.

In this section, the solutions encountered in the two steps of the algorithm will be analyzed and the usability of an iterative solution in lieu of or complementing an analytical solution assessed. In doing so, the derivations due to Section 3.6 will be of sufficient aid and will be the main reference.

First, note from Appendix B.3 and relations (3.48) and (3.49) that the E-step of the EM algorithm is simply

$$\langle \mathbf{x} \rangle_{i,l} = \mathbf{\Omega}_i \mathbf{W}_i^* (\mathsf{S}_i^{\mathsf{z}})^{-1} \mathbf{y}(\omega_l) \tag{5.20}$$

$$\mathbf{\Omega}_i = (I_q + \mathbf{W}_i^* (\mathsf{S}_i^{\mathsf{z}})^{-1} \mathbf{W}_i)^{-1}, \tag{5.21}$$

$$\langle \mathbf{x}\mathbf{x}^* \rangle_{i,l} = \langle \mathbf{x} \rangle_{i,l} \langle \mathbf{x} \rangle_{i,l}^* + \mathbf{\Omega}_i, \tag{5.22}$$

where the inverse of $\mathbf{\Omega}_i \in \mathbb{C}^{q \times q}$, in general, exists. For evaluating $\mathbf{W}_{i+1}$ according to (5.18), first write

$$\begin{aligned} \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})] &= \sum_{l=1}^{n} \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\log_e(\mathcal{N}_{\mathbb{C}}(\mathbf{y}(\omega_l) \mid \mathbf{W}_i \mathsf{x}, \mathsf{S}_i^{\mathsf{z}}))] \\ &= -n \log_e(|\mathsf{S}_i^{\mathsf{z}}|) \\ &\quad - \sum_{l=1}^{n} \mathrm{tr}((\mathsf{S}_i^{\mathsf{z}})^{-1} \mathbf{W}_i \langle \mathbf{x}\mathbf{x}^* \rangle_{i,l} \mathbf{W}_i^*) + \mathbf{y}^*(\omega_l)(\mathsf{S}_i^{\mathsf{z}})^{-1}\mathbf{y}(\omega_l) \\ &\qquad - 2\Re(\mathbf{y}^*(\omega_l)(\mathsf{S}_i^{\mathsf{z}})^{-1} \mathbf{W}_i \langle \mathbf{x} \rangle_{i,l}), \end{aligned} \tag{5.23}$$

where eliminated are terms independent of $\mathbf{W}_i$ or $\mathsf{S}_i^{\mathsf{z}}$. The reader is referred to [57] to verify using Wirtinger relaxations that

$$\frac{\partial}{\partial \mathbf{W}_i} \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})] = (\mathsf{S}_i^{\mathsf{z}})^{-1} \sum_{l=1}^{n} (\overline{\mathbf{W}}_i \langle \mathbf{x}\mathbf{x}^* \rangle_{i,l}' - \overline{\mathbf{y}}(\omega_l) \langle \mathbf{x} \rangle_{i,l}'). \tag{5.24}$$

Then, due to (5.18),

$$\mathbf{W}_{i+1} = \left(\sum_{l=1}^{n} \mathbf{y}(\omega_l)(\langle \mathbf{x} \rangle_{i,l})^* \right) \left(\sum_{l=1}^{n} \langle \mathbf{x}\mathbf{x}^* \rangle_{i,l} \right)^{-1}. \tag{5.25}$$

Just as in Section 3.7.2, let

$$\mathbf{v}_i(\omega_l) \triangleq \mathbf{W}_{i+1} \langle \mathbf{x} \rangle_{i,l}. \tag{5.26}$$

For $\mathsf{S}_i^{\mathsf{z}} = \mathrm{diag}(\mathsf{s}_i^{\mathsf{z}_1}, \ldots, \mathsf{s}_i^{\mathsf{z}_r})$ it can easily be seen that

$$\mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})] = -n \sum_{k=1}^{r} [\log_e(\mathsf{s}_i^{\mathsf{z}_k}) + \frac{1}{\mathsf{s}_i^{\mathsf{z}_k}} b_i^{\mathsf{z}_k}] \tag{5.27}$$

where

$$b_i^{z_k} = \frac{1}{n}\sum_{l=1}^{n} |\mathbf{y}_k(\omega_l) - \mathbf{v}_{ki}(\omega_l)|^2. \tag{5.28}$$

Note that $\mathsf{S}_i^{\mathsf{z}}$ is a real-valued diagonal matrix and the derivative with respect to it is straightforward. Then $\partial \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta_i, \mathsf{x})]/\partial \mathsf{s}_i^{\mathsf{z}_k} = 0$ at

$$\begin{aligned} \mathsf{s}_{i+1}^{\mathsf{z}_k} &= b_i^{\mathsf{z}_k}, \\ \mathsf{S}_{i+1}^{\mathsf{z}} &= \mathsf{diag}(\mathsf{s}_{i+1}^{\mathsf{z}_1}, \cdots, \mathsf{s}_{i+1}^{\mathsf{z}_r}). \end{aligned} \tag{5.29}$$

The relations (5.25) and (5.29) stand for the M-step of the EM algorithm for the maximum likelihood parameters of the spectral factor model.

5.2.2 EM algorithm for spectral factor model

The following pseudocode of the EM algorithm for the maximum likelihood spectral factor model may now be provided; this is in line with Algorithm 2 in Section 3.7.2. In Algorithm 2, the input was the dataset $\mathcal{D}$ of iid data samples; whereas here it is assumed that $\mathcal{D}$ is a set of discrete Fourier transform components near a target frequency as recommended by the asymptotic requirements of Theorem 2.5.

Algorithm 4: EM algorithm for the spectral factor model in a subband

```
Input: D = {y(ω_l)}, l = 1, ... n
Output: Ŵ, Ŝ^z = diag(ŝ^{z_1}, ..., ŝ^{z_r})
initialize i = 0;
randomize W_i, S_i^z;
do
    E-step:
    for l = 1 to n do
        compute
            ⟨x⟩_{i,l} using (5.20);
            ⟨xx*⟩_{i,l} using (5.22);
    end
    M-step: update
            W_{i+1} using (5.25);
            s_{i+1}^{z_k} ∀k = 1, ..., r using (5.29);
    i ⟵ i + 1;
    ϵ ⟵ E^{x|y,θ}[f(θ_i, x)] − E^{x|y,θ}[f(θ_{i−1}, x)] using (5.23);
while ϵ > 10^{−8} and i < 20;
Ŵ ⟵ W_i, ŝ^{z_k} ⟵ s_i^{z_k} ∀k = 1, ..., r;
```

Suppose the discrete Fourier transform components $\mathcal{D} = \{\mathbf{y}(\omega_l)\}$, $l = 1, \ldots n$ within a subband are obtained as per Algorithm 1. Algorithm 4 demonstrates how the E and M steps may be alternated, starting with a random initialization of the parameters corresponding to a target frequency, to output the converged parameters $\widehat{\mathbf{W}}$ and $\widehat{\mathsf{S}}^{\mathsf{z}}$ of the spectral factor model.

Note 5.4. *For EM algorithm in Algorithm 4 converging towards a local maximum of the log-likelihood is possible, converged parameters θ corresponding to the largest* $\mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{f}(\theta,\mathsf{x})]$ *from a number of random restarts will be chosen.*

5.2.3 Maximizing commonalities in spectral factor model

Note 5.5. *In this section, it is assumed that the EM steps have converged. Therefore, for notational brevity, any indexing of the iteration is dropped and the updated parameters will be denoted by $\theta \triangleq \{\widehat{\mathbf{W}}, \widehat{\mathsf{S}}^{\mathsf{z}}\}$. As in the previous sections, ω_l corresponds to the l-th frequency sample within the subband near the target frequency.*

As seen, at the end of the iterations of a converged EM algorithm, access is available to the estimate of the transformed factor to get

$$\mathbf{v}(\omega_l) = \widehat{\mathbf{W}}\mathbf{x}(\omega_l)$$

corresponding to the l-th realization $\mathbf{y}(\omega_l)$ $\forall l \in 1,\ldots,n$, where

$$\mathbf{x}(\omega_l) = \mathsf{E}^{\mathsf{x}|\mathsf{y},\theta}[\mathsf{x} \mid \mathbf{y}(\omega_l), \theta] \tag{5.30}$$

as in (5.17) and computed in (5.20). Thus, in the context of (4.15),

$$\mathbf{y}(\omega_l) = \widehat{\mathbf{W}}\mathbf{x}(\omega_l) + \mathbf{z}(\omega_l) \quad \forall l = 1,\ldots,n \tag{5.31}$$

where $\mathbf{z}(\omega_l)$ is the error in regressing $\mathbf{x}(\omega_l)$ towards $\mathbf{y}(\omega_l)$. The regression errors are due to zero mean isotropic Gaussian vector random variable z; this is not the assumption but the result of Theorem 2.5.

Now it shall be seen why the same situation as in the linear model of Section 3.2 persists. From the form of (4.18) for inheritance by S^{v} of the commonalities of S^{y}, it is clear that $\|\mathsf{S}^{\mathsf{y}}(\omega) - \mathsf{S}^{\mathsf{v}}(\omega)\|_F^2$ may be minimized for each ω individually. Then, (4.19) implies that the optimal S^{z} is given by $\widetilde{\mathsf{S}}^{\mathsf{z}} \triangleq \underset{\mathsf{S}^{\mathsf{z}}}{\operatorname{argmin}}\, g = \underset{\mathsf{S}^{\mathsf{z}}}{\operatorname{argmin}}\, \operatorname{tr}(\mathsf{S}^{\mathsf{z}})$, or for each of the diagonal elements $\mathsf{s}^{\mathsf{z}_k}$ of S^{z}

$$\tilde{\mathsf{s}}^{\mathsf{z}_k} \triangleq \min(\mathsf{s}^{\mathsf{z}_k}) \quad \forall k = 1,\ldots,r. \tag{5.32}$$

But $\mathsf{s}^{\mathsf{z}_k}$ is the variance of the zero mean Gaussian error in approximating $\mathbf{y}_k(\omega_l)$ using $\mathbf{v}_k(\omega_l)$. Hence, a minimum variance unbiased regression of $\mathbf{x}(\omega_l)$ towards $\mathbf{y}(\omega_l)$ is sought using $\mathbf{v}(\omega_l) = \mathbf{W}\mathbf{x}(\omega_l)$.

As seen, maximizing the commonalities upon convergence of the EM algorithm requires an efficient estimator of $\mathbf{W}$. Therefore, if the Gauss-Markov solution of (3.8) is used, the efficient estimator got is

$$\widetilde{\mathbf{W}} = [\mathbf{y}(\omega_1)\cdots\mathbf{y}(\omega_n)]\mathbf{X}^*\left(\mathbf{X}\mathbf{X}^*\right)^{-1}, \tag{5.33}$$

where, using $\mathbf{x}(\omega_l)$ referred to in (5.30) and computed via (5.20), the $q \times n$ matrix $\mathbf{X} = [\mathbf{x}(\omega_1) \cdots \mathbf{x}(\omega_n)]$ having rank($\mathbf{X}$) $= q$ is a maximum likelihood 'latent data matrix.' And, as per (3.9), an unbiased estimate of S^{z} is

$$\begin{aligned} \widetilde{\mathsf{S}}^{\mathsf{z}} &= \mathrm{diag}(\tilde{\mathsf{s}}^{\mathsf{z}_1}, \ldots, \tilde{\mathsf{s}}^{\mathsf{z}_r}), \\ \tilde{\mathsf{s}}^{\mathsf{z}_k} &= \frac{1}{n-q} \sum_{l=1}^{n} |\mathbf{y}_k(\omega_l) - \tilde{\mathbf{v}}_k(\omega_l)|^2, \; k = 1, \ldots, r, \\ \tilde{\mathbf{v}}(\omega_l) &= \widetilde{\mathbf{W}} \mathbf{x}(\omega_l), \end{aligned} \tag{5.34}$$

where $\mathbf{x}(\omega_l)$ referred to in (5.30) is computed via (5.20). It is important to understand that although the EM algorithm gives the maximum likelihood solution, the maximization of the commonalities was achieved through (5.33).

Suppose the discrete Fourier transform components $\mathcal{D} = \{\mathbf{y}(\omega_l)\}, l = 1, \ldots n$ within a subband as per Algorithm 1 are obtained. Then, as per Algorithm 5, the procedure for estimating the maximum commonalities spectral factor model parameters utilizing the EM algorithm developed in Algorithm 4 could be compiled.

Algorithm 5: Maximum commonalities spectral factor model via EM algorithm

Input: $\mathcal{D} = \{\mathbf{y}(\omega_l)\}, l = 1, \ldots n$
Output: $\widetilde{\mathbf{W}}, \widetilde{\mathsf{S}}^{\mathsf{z}}$
estimate $\{\widehat{\mathbf{W}}, \widehat{\mathsf{S}}^{z}\}$ with input $\mathcal{D}$ to Algorithm 4;
compute $\mathbf{x}(\omega_l)$ as in (5.30);
estimate $\widetilde{\mathbf{W}}$ using (5.33);
estimate $\widetilde{\mathsf{S}}^{\mathsf{z}}$ using (5.34);

5.3 Summary

The form of the spectral factor model in (4.15) is similar to the classical factor model in (3.11). In this chapter, as reviewed for the classical factor model in Chapter 3, two approaches for maximum likelihood estimation of the spectral factor model were developed and within each of them the commonality maximization parameters were found:

In the analytical approach put forth, the sample spectral density function computed from the discrete Fourier transform samples of a measured vector random process near a target frequency is the maximum likelihood spectral density function of the process. The maximum likelihood maximum commonalities solution provided by (5.10) is similar in interpretation to the low-rank approximation of the classical factor model solution and (5.13) provides the leverage to choose idiosyncrasies the way the user wants without destroying the rank stringencies of the transformation matrix. The commonality maximizing maximum likelihood transformation was found to direct the latent spectra along the principal components of the measured spectra. This analytical solution was presented in Algorithm 3.

Again, as with the classical factor model, Algorithm 4 was designed to estimate the maximum likelihood spectral factor model in an iterative fashion. The parameters of the model thus estimated were improved in (5.33) and (5.34) by treating the maximum likelihood transformation of the *a posteriori* mean of the latent variables of the spectral model as a regression towards the measured spectral components. This enabled the transformed latent spectra to maximally inherit the commonalities of the measured spectra through the Gauss-Markov theorem.