Implementation details of OSC-DECOMP

March 2022
Takeru Matsuda
takeru.matsuda@riken.jp

This is a note of implementaion details of OSC-DECOMP. Most of them are not described in the original papers [4, 5].

1 Reparametrization

The state space model of OSC-DECOMP is

$$x_{t+1} = Fx_t + v_t, \quad v_t \sim \mathrm{N}(0, Q),$$
$$y_t = Hx_t + w_t, \quad w_t \sim \mathrm{N}(0, R),$$

where

$$F = \begin{pmatrix} F_1 & O & \cdots & O \\ O & F_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & F_K \end{pmatrix}, \quad F_k = a_k \begin{pmatrix} \cos(2\pi f_k \Delta t) & -\sin(2\pi f_k \Delta t) \\ \sin(2\pi f_k \Delta t) & \cos(2\pi f_k \Delta t) \end{pmatrix}, \tag{1}$$

$$Q = \begin{pmatrix} Q_1 & O & \cdots & O \\ O & Q_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & Q_K \end{pmatrix}, \quad Q_k = \begin{pmatrix} \sigma_k^2 & 0 \\ 0 & \sigma_k^2 \end{pmatrix}, \tag{2}$$

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & \cdots & 1 & 0 \\ c_{21,1} & c_{21,2} & c_{22,1} & c_{22,2} & \cdots & c_{2K,1} & c_{2K,2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \\ c_{J1,1} & c_{J1,2} & c_{J2,1} & c_{J2,2} & \cdots & c_{JK,1} & c_{JK,2} \end{pmatrix}, \quad R = \tau^2 I_J. \tag{3}$$

The parameter is $\theta = (a_1, \dots, a_K, f_1, \dots, f_K, \sigma_1^2, \dots, \sigma_K^2, c_{21,1}, \dots, c_{JK,2}, \tau^2)$, where $0 < a_k < 1$, $0 \leq 2\pi f_k \Delta t \leq \pi$, $\sigma_k^2 > 0$ and $\tau^2 > 0$.

To formulate the model fitting as an unconstrained optimization of the log-likelihood, the parameter θ is bijectively transformed as follows:

- $a_k \in (0,1)$ is transformed to $\tilde{a}_k = \tanh^{-1}(2a_k - 1) \in (-\infty, \infty)$ so that $a_k = (\tanh(\tilde{a}_k) + 1)/2$.
- $f_k \in [0, (2\Delta t)^{-1}]$ is transformed to $\tilde{f}_k = \tanh^{-1}(2(f_k - f_{k0})\Delta t) \in (-\infty, \infty)$ so that $2\pi f_k \Delta t = 2\pi f_{k0} \Delta t + \pi \tanh(\tilde{f}_k)$, where $(f_k)_0$ is the initial guess of f_k (see Appendix of [4]). Thus, the initial value of ξ_k is zero. This parameter transformation avoids the numerical instability due to the circular nature of frequency variable coming from the aliasing effect.
- $\sigma_k^2 > 0$ is transformed to $\log(\sigma_k^2/\tau^2) \in (-\infty, \infty)$.
- $c_{jk,l} \in (-\infty, \infty)$ is already unconstrained and thus kept unchanged.
- $\tau^2 > 0$ is kept unchanged and optimized in closed form (see Section 2).

2 Optimization algorithm

Parameter estimation is carried out by maximizing the log-likelihood $\log L(\widetilde{\theta})$, where $\widetilde{\theta}$ is the transformed parameter introduced in the Section 1. The maximization of $\log L(\widetilde{\theta})$ with respect to the observation noise variance τ^2 can be solved in closed form. See [4, 5] for its detail. Thus, the problem is reduced to the maximization of the profile log-likelihood $\max_{\tau^2} \log L(\widetilde{\theta})$ with respect to the remaining elements of $\widetilde{\theta}$.

We use the MATLAB function "fminunc" for unconstrained optimization in maximizing the profile log-likelihood. This function implements the quasi-Newton and trust-region algorithms[1]. In the quasi-Newton algorithm, we can choose whether to provide the gradient explicitly or rely on numerical differentiation. In the trust-region algorithm, the gradient is always required.

Tables 1 and 2 compare the performance of the optimization algorithms for Canadian Lynx data and north-south sunspot data analyzed in [4, 5], respectively[2]. For Canadian Lynx data, three algorithms converge to parameter estimates with almost the same AIC values. For north-south sunspot data, quasi-Newton with numerical differentiaton attains smaller AIC value than the others. In terms of the computational time, quasi-Newton with numerical differentiaton is the fastest. Thus, the cost of gradient computation by the Kalman filter (see Section 3) seems to overwhelm that of numerical differentiation. Also, numerical experiments in [3] show that the accuracy of gradient computation by the Kalman filter may be limited. Thus, we set the quasi-Newton with numerical differentiaton to the default optimization algorithm. If necessary, other procedures can be used by setting the arguments "algorithm" and "grad" appropriately.

Table 1: Result on Canadian Lynx data (MAX_OSC=6)

	time (seconds)	minAIC
quasi-Newton (numerical diff.)	16.70	166.38 ($K = 4$)
quasi-Newton (gradient by KF)	26.30	166.38 ($K = 4$)
trust-region (gradient by KF)	106.80	166.40 ($K = 4$)

Table 2: Result on north-south sunspot data (MAX_OSC=6)

	time (seconds)	minAIC
quasi-Newton (numerical diff.)	275.34	955.14 ($K = 5$)
quasi-Newton (gradient by KF)	411.52	955.21 ($K = 5$)
trust-region (gradient by KF)	773.69	955.14 ($K = 5$)

We also tried the nonlinear conjugate gradient method [7], but it did not converge well for some reason. Another possible choice is the EM algorithm [8], although it may not be fast compared to numerical optimization.

Since the objective function is non-convex, starting from a good initial guess is also crucial for obtaining a good parameter estimate. The current method for initial guess setting based on AR/VAR models [4, 5] may be improved and it is an important future (ongoing) problem.

[1] The information here is based on MATLAB2020b.
[2] script: opt_cmp_uni.m, opt_cmp_multi.m.

3 Gradient and Hessian computation

The gradient and Hessian of the log-likelihood can be computed by an extension of the Kalman filter algorithm. It is a generalization of the algorithm in [3] to multi-dimensional observations.

Consider a Gaussian linear state space model in general form:

$$\begin{aligned} x_{t+1} &= F(\theta)x_t + v_t, \quad v_t \sim \mathrm{N}_n(0, Q(\theta)), \\ y_t &= H(\theta)x_t + w_t, \quad w_t \sim \mathrm{N}_m(0, R(\theta)), \end{aligned}$$

where $x_t \in \mathbb{R}^n$, $y_t \in \mathbb{R}^m$, $F \in \mathbb{R}^{n\times n}$, and $H \in \mathbb{R}^{m\times n}$. In the following, we omit the argument θ for simplicity. The one-step ahead predictive distribution of y_t is

$$p(y_t \mid y_1, \ldots, y_{t-1}, \theta) = \frac{1}{(2\pi)^{m/2}(\det R_t)^{1/2}} \exp\left(-\frac{1}{2}\varepsilon_t^\top R_t^{-1}\varepsilon_t\right),$$

where ε_t and R_t are the one-step ahead prediction error and its covariance defined by

$$\varepsilon_t = y_t - Hx_{t|t-1}, \quad R_t = H\Sigma_{t|t-1}H^\top + R. \tag{4}$$

Therefore, the log-likelihood is given by

$$\begin{aligned} \log L(\theta) &= \sum_{t=1}^{N} \log p(y_t \mid y_1, \ldots, y_{t-1}, \theta) \\ &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{N}(\log\det R_t + \varepsilon_t^\top R_t^{-1}\varepsilon_t). \end{aligned} \tag{5}$$

By differentiating (5), the gradient of the log-likelihood is

$$\frac{\partial}{\partial\theta_i}\log L(\theta) = -\frac{1}{2}\sum_{t=1}^{N}\left(\mathrm{tr}\left(R_t^{-1}\frac{\partial R_t}{\partial\theta_i}\right) + 2\varepsilon_t^\top R_t^{-1}\frac{\partial\varepsilon_t}{\partial\theta_i} - \varepsilon_t^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_i}R_t^{-1}\varepsilon_t\right), \tag{6}$$

where, from (4),

$$\frac{\partial\varepsilon_t}{\partial\theta_i} = -\frac{\partial H}{\partial\theta_i}x_{t|t-1} - H\frac{\partial x_{t|t-1}}{\partial\theta_i}, \tag{7}$$

$$\frac{\partial R_t}{\partial\theta_i} = \frac{\partial H}{\partial\theta_i}\Sigma_{t|t-1}H^\top + H\frac{\partial\Sigma_{t|t-1}}{\partial\theta_i}H^\top + H\Sigma_{t|t-1}\frac{\partial H^\top}{\partial\theta_i} + \frac{\partial R}{\partial\theta_i}. \tag{8}$$

From the Kalman filter formula, the required derivatives are computed sequentially as

$$\frac{\partial x_{t|t-1}}{\partial\theta_i} = \frac{\partial F}{\partial\theta_i}x_{t-1|t-1} + F\frac{\partial x_{t-1|t-1}}{\partial\theta_i}, \tag{9}$$

$$\frac{\partial\Sigma_{t|t-1}}{\partial\theta_i} = \frac{\partial F}{\partial\theta_i}\Sigma_{t-1|t-1}F^\top + F\frac{\partial\Sigma_{t-1|t-1}}{\partial\theta_i}F^\top + F\Sigma_{t-1|t-1}\frac{\partial F^\top}{\partial\theta_i} + \frac{\partial Q}{\partial\theta_i}, \tag{10}$$

and

$$\frac{\partial K_t}{\partial\theta_i} = \frac{\partial\Sigma_{t|t-1}}{\partial\theta_i}H^\top R_t^{-1} + \Sigma_{t|t-1}\frac{\partial H^\top}{\partial\theta_i}R_t^{-1} - \Sigma_{t|t-1}H^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_i}R_t^{-1}, \tag{11}$$

$$\frac{\partial x_{t|t}}{\partial\theta_i} = \frac{\partial x_{t|t-1}}{\partial\theta_i} + \frac{\partial K_t}{\partial\theta_i}\varepsilon_t + K_t\frac{\partial\varepsilon_t}{\partial\theta_i}, \tag{12}$$

$$\frac{\partial\Sigma_{t|t}}{\partial\theta_i} = \frac{\partial\Sigma_{t|t-1}}{\partial\theta_i} - \frac{\partial K_t}{\partial\theta_i}H\Sigma_{t|t-1} - K_t\frac{\partial H}{\partial\theta_i}\Sigma_{t|t-1} - K_tH\frac{\partial\Sigma_{t|t-1}}{\partial\theta_i}, \tag{13}$$

where $K_t = \Sigma_{t|t-1} H^\top R_t^{-1}$ is the Kalman gain.

By differentiating (6) and using

$$\frac{\partial}{\partial \theta_i}(R_t^{-1}) = -R_t^{-1} \frac{\partial R_t}{\partial \theta_i} R_t^{-1},$$

the Hessian of the log-likelihood is

$$\begin{aligned}
\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\theta) = -\frac{1}{2} \sum_{t=1}^{N} \Bigg(& \operatorname{tr}\left(-R_t^{-1} \frac{\partial R_t}{\partial \theta_j} R_t^{-1} \frac{\partial^2 R_t}{\partial \theta_i} + R_t^{-1} \frac{\partial^2 R_t}{\partial \theta_i \partial \theta_j} \right) \\
& + 2 \frac{\partial \varepsilon_t^\top}{\partial \theta_j} R_t^{-1} \frac{\partial \varepsilon_t}{\partial \theta_i} - 2\varepsilon_t^\top R_t^{-1} \frac{\partial R_t}{\partial \theta_j} R_t^{-1} \frac{\partial \varepsilon_t}{\partial \theta_i} + 2\varepsilon_t^\top R_t^{-1} \frac{\partial^2 \varepsilon_t}{\partial \theta_i \partial \theta_j} \\
& - 2 \frac{\partial \varepsilon_t^\top}{\partial \theta_j} R_t^{-1} \frac{\partial R_t}{\partial \theta_i} R_t^{-1} \varepsilon_t + 2\varepsilon_t^\top R_t^{-1} \frac{\partial R_t}{\partial \theta_j} R_t^{-1} \frac{\partial R_t}{\partial \theta_i} R_t^{-1} \varepsilon_t \\
& - \varepsilon_t^\top R_t^{-1} \frac{\partial^2 R_t}{\partial \theta_i \partial \theta_j} R_t^{-1} \varepsilon_t \Bigg),
\end{aligned}$$

where, from (8),

$$\begin{aligned}
\frac{\partial^2 \varepsilon_t}{\partial \theta_i \partial \theta_j} =& - \frac{\partial^2 H}{\partial \theta_i \partial \theta_j} x_{t|t-1} - \frac{\partial H}{\partial \theta_i} \frac{\partial x_{t|t-1}}{\partial \theta_j} - \frac{\partial H}{\partial \theta_j} \frac{\partial x_{t|t-1}}{\partial \theta_i} - H \frac{\partial^2 x_{t|t-1}}{\partial \theta_i \partial \theta_j}, \\
\frac{\partial^2 R_t}{\partial \theta_i \partial \theta_j} =& \frac{\partial^2 H}{\partial \theta_i \partial \theta_j} \Sigma_{t|t-1} H^\top + \frac{\partial H}{\partial \theta_i} \frac{\partial \Sigma_{t|t-1}}{\partial \theta_j} H^\top + \frac{\partial H}{\partial \theta_i} \Sigma_{t|t-1} \frac{\partial H^\top}{\partial \theta_j} \\
& + \frac{\partial H}{\partial \theta_j} \frac{\partial \Sigma_{t|t-1}}{\partial \theta_i} H^\top + H \frac{\partial^2 \Sigma_{t|t-1}}{\partial \theta_i \partial \theta_j} H^\top + H \frac{\partial \Sigma_{t|t-1}}{\partial \theta_i} \frac{\partial H^\top}{\partial \theta_j} \\
& + \frac{\partial H}{\partial \theta_j} \Sigma_{t|t-1} \frac{\partial H^\top}{\partial \theta_i} + H \frac{\partial \Sigma_{t|t-1}}{\partial \theta_j} \frac{\partial H^\top}{\partial \theta_i} + H \Sigma_{t|t-1} \frac{\partial^2 H^\top}{\partial \theta_i \partial \theta_j} \\
& + \frac{\partial^2 R}{\partial \theta_i \partial \theta_j}.
\end{aligned}$$

From (9)-(13), the required derivatives are computed sequentially as

$$\begin{aligned}
\frac{\partial^2 x_{t|t-1}}{\partial \theta_i \partial \theta_j} =& \frac{\partial^2 F}{\partial \theta_i \partial \theta_j} x_{t-1|t-1} + \frac{\partial F}{\partial \theta_i} \frac{\partial x_{t-1|t-1}}{\partial \theta_j} \\
& + \frac{\partial F}{\partial \theta_j} \frac{\partial x_{t-1|t-1}}{\partial \theta_i} + F \frac{\partial^2 x_{t-1|t-1}}{\partial \theta_i \partial \theta_j}, \\
\frac{\partial^2 \Sigma_{t|t-1}}{\partial \theta_i \partial \theta_j} =& \frac{\partial^2 F}{\partial \theta_i \partial \theta_j} \Sigma_{t-1|t-1} F^\top + \frac{\partial F}{\partial \theta_i} \frac{\partial \Sigma_{t-1|t-1}}{\partial \theta_j} F^\top + \frac{\partial F}{\partial \theta_i} \Sigma_{t-1|t-1} \frac{\partial F^\top}{\partial \theta_j} \\
& + \frac{\partial F}{\partial \theta_j} \frac{\partial \Sigma_{t-1|t-1}}{\partial \theta_i} F^\top + F \frac{\partial^2 \Sigma_{t-1|t-1}}{\partial \theta_i \partial \theta_j} F^\top + F \frac{\partial \Sigma_{t-1|t-1}}{\partial \theta_i} \frac{\partial F^\top}{\partial \theta_j} \\
& + \frac{\partial F}{\partial \theta_j} \Sigma_{t-1|t-1} \frac{\partial F^\top}{\partial \theta_i} + F \frac{\partial \Sigma_{t-1|t-1}}{\partial \theta_j} \frac{\partial F^\top}{\partial \theta_i} + F \Sigma_{t-1|t-1} \frac{\partial^2 F^\top}{\partial \theta_i \partial \theta_j} \\
& + \frac{\partial^2 Q}{\partial \theta_i \partial \theta_j},
\end{aligned}$$

and

$$
\begin{aligned}
\frac{\partial^2 K_t}{\partial\theta_i\partial\theta_j} =& \frac{\partial^2 \Sigma_{t|t-1}}{\partial\theta_i\partial\theta_j} H^\top R_t^{-1} + \frac{\partial \Sigma_{t|t-1}}{\partial\theta_i}\frac{\partial H^\top}{\partial\theta_j} R_t^{-1} - \frac{\partial \Sigma_{t|t-1}}{\partial\theta_i} H^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_j} R_t^{-1} \\
&+ \frac{\partial \Sigma_{t|t-1}}{\partial\theta_j}\frac{\partial H^\top}{\partial\theta_i} R_t^{-1} + \Sigma_{t|t-1}\frac{\partial^2 H^\top}{\partial\theta_i\partial\theta_j} R_t^{-1} - \Sigma_{t|t-1}\frac{\partial H^\top}{\partial\theta_i} R_t^{-1}\frac{\partial R_t}{\partial\theta_j} R_t^{-1} \\
&- \frac{\partial \Sigma_{t|t-1}}{\partial\theta_j} H^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_i} R_t^{-1} - \Sigma_{t|t-1}\frac{\partial H^\top}{\partial\theta_j} R_t^{-1}\frac{\partial R_t}{\partial\theta_i} R_t^{-1} \\
&+ \Sigma_{t|t-1} H^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_j} R_t^{-1}\frac{\partial R_t}{\partial\theta_i} R_t^{-1} - \Sigma_{t|t-1} H^\top R_t^{-1}\frac{\partial^2 R_t}{\partial\theta_i\partial\theta_j} R_t^{-1} \\
&+ \Sigma_{t|t-1} H^\top R_t^{-1}\frac{\partial R_t}{\partial\theta_i} R_t^{-1}\frac{\partial R_t}{\partial\theta_j} R_t^{-1}, \\
\frac{\partial^2 x_{t|t}}{\partial\theta_i\partial\theta_j} =& \frac{\partial^2 x_{t|t-1}}{\partial\theta_i\partial\theta_j} + \frac{\partial^2 K_t}{\partial\theta_i\partial\theta_j}\varepsilon_t + \frac{\partial K_t}{\partial\theta_i}\frac{\partial \varepsilon_t}{\partial\theta_j} + \frac{\partial K_t}{\partial\theta_j}\frac{\partial \varepsilon_t}{\partial\theta_i} + K_t\frac{\partial^2 \varepsilon_t}{\partial\theta_i\partial\theta_j} \\
\frac{\partial^2 \Sigma_{t|t}}{\partial\theta_i\partial\theta_j} =& \frac{\partial^2 \Sigma_{t|t-1}}{\partial\theta_i\partial\theta_j} - \frac{\partial^2 K_t}{\partial\theta_i\partial\theta_j} H\Sigma_{t|t-1} - \frac{\partial K_t}{\partial\theta_i}\frac{\partial H}{\partial\theta_j}\Sigma_{t|t-1} - \frac{\partial K_t}{\partial\theta_i} H\frac{\partial \Sigma_{t|t-1}}{\partial\theta_j} \\
&- \frac{\partial K_t}{\partial\theta_j}\frac{\partial H}{\partial\theta_i}\Sigma_{t|t-1} - K_t\frac{\partial^2 H}{\partial\theta_i\partial\theta_j}\Sigma_{t|t-1} - K_t\frac{\partial H}{\partial\theta_i}\frac{\partial \Sigma_{t|t-1}}{\partial\theta_j} \\
&- \frac{\partial K_t}{\partial\theta_j} H\frac{\partial \Sigma_{t|t-1}}{\partial\theta_i} - K_t\frac{\partial H}{\partial\theta_j}\frac{\partial \Sigma_{t|t-1}}{\partial\theta_i} - K_t H\frac{\partial^2 \Sigma_{t|t-1}}{\partial\theta\partial\theta_i^\top}.
\end{aligned}
$$

Now, we provide required derivatives for the state space model in OSC-DECOMP. From (1), (2) and (3), the nonzero derivatives of F, Q, H, R are

$$
\begin{aligned}
\frac{\partial F_k}{\partial a_k} &= \begin{pmatrix} \cos(2\pi f_k\Delta t) & -\sin(2\pi f_k\Delta t) \\ \sin(2\pi f_k\Delta t) & \cos(2\pi f_k\Delta t) \end{pmatrix}, \\
\frac{\partial F_k}{\partial f_k} &= 2\pi a_k\Delta t\begin{pmatrix} -\sin(2\pi f_k\Delta t) & -\cos(2\pi f_k\Delta t) \\ \cos(2\pi f_k\Delta t) & -\sin(2\pi f_k\Delta t) \end{pmatrix}, \\
\frac{\partial^2 F_k}{\partial a_k\partial f_k} &= 2\pi\Delta t\begin{pmatrix} -\sin(2\pi f_k\Delta t) & -\cos(2\pi f_k\Delta t) \\ \cos(2\pi f_k\Delta t) & -\sin(2\pi f_k\Delta t) \end{pmatrix}, \\
\frac{\partial^2 F_k}{\partial f_k^2} &= 4\pi^2 a_k(\Delta t)^2\begin{pmatrix} -\cos(2\pi f_k\Delta t) & \sin(2\pi f_k\Delta t) \\ -\sin(2\pi f_k\Delta t) & -\cos(2\pi f_k\Delta t) \end{pmatrix}, \\
\frac{\partial Q_k}{\partial(\sigma_k^2)} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \frac{\partial H_{j,2k+l-2}}{\partial c_{jk,l}} = 1, \quad \frac{\partial R}{\partial(\tau^2)} = I_J.
\end{aligned}
$$

Also, from the stationary distribution

$$
x_{1|0} = 0, \quad (V_{1|0})_{2k-1,2k-1} = (V_{1|0})_{2k,2k} = \frac{\sigma_k^2}{1-a_k^2}, \quad (V_{1|0})_{ij} = 0 \ (i \neq j),
$$

the initial gradient and Hessian are

$$
\begin{aligned}
&\frac{\partial x_{1|0}}{\partial \theta_i} = 0, \quad \frac{\partial^2 x_{1|0}}{\partial \theta_i \partial \theta_j} = 0, \\
&\frac{\partial (V_{1|0})_{2k-1,2k-1}}{\partial a_k} = \frac{\partial (V_{1|0})_{2k,2k}}{\partial a_k} = \frac{2a_k \sigma_k^2}{(1-a_k^2)^2}, \\
&\frac{\partial (V_{1|0})_{2k-1,2k-1}}{\partial (\sigma_k^2)} = \frac{\partial (V_{1|0})_{2k,2k}}{\partial (\sigma_k^2)} = \frac{1}{1-a_k^2}, \\
&\frac{\partial^2 (V_{1|0})_{2k-1,2k-1}}{\partial a_k^2} = \frac{\partial^2 (V_{1|0})_{2k,2k}}{\partial a_k^2} = \frac{2(3a_k^2+1)\sigma_k^2}{(1-a_k^2)^3}, \\
&\frac{\partial^2 (V_{1|0})_{2k-1,2k-1}}{\partial a_k \partial (\sigma_k^2)} = \frac{\partial^2 (V_{1|0})_{2k,2k}}{\partial a_k \partial (\sigma_k^2)} = \frac{2a_k}{(1-a_k^2)^2},
\end{aligned}
$$

where other derivatives of $V_{1|0}$ are zero.

4 Confidence interval

Let

$$
H(\theta) = -\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} \log L(\tilde{\theta}) \right)_{ij}
$$

be the Hessian of the negative log-likelihood. For the maximum likelihood estimate $\hat{\theta}$, the matrix $H(\hat{\theta})$ is often called the observed Fisher information [2]. From the theory of asymptotic statistics [?], the asymptotic covariance matrix of $\hat{\theta}$ can be approximated by $V = H(\hat{\theta})^{-1}$. Therefore, the 95 % confidence interval of θ_i can be constructed as

$$
\left[\hat{\theta}_i - 1.96\sqrt{V_{ii}}, \hat{\theta}_i + 1.96\sqrt{V_{ii}} \right].
$$

For the phase difference $\phi_{jk} = \arg(c_{jk,1} + c_{jk,2}\sqrt{-1})$, its confidence interval can be computed by using the delta method [6]. Let $V_{(jk)}$ be the 2×2 submarix of V corresponding to $c_{jk,1}$ and $c_{jk,2}$. Then, from

$$
\frac{\partial \phi_{jk}}{\partial c_{jk,1}} = -\frac{c_{jk,2}}{c_{jk,1}^2 + c_{jk,2}^2}, \quad \frac{\partial \phi_{jk}}{\partial c_{jk,2}} = \frac{c_{jk,1}}{c_{jk,1}^2 + c_{jk,2}^2},
$$

the asymptotic variance of $\hat{\phi}_{jk} = \arg(\hat{c}_{jk,1} + \hat{c}_{jk,2}\sqrt{-1})$ is approximated by

$$
v_{(jk)} = \frac{\hat{c}_{jk,2}^2}{(\hat{c}_{jk,1}^2 + \hat{c}_{jk,2}^2)^2} (V_{(jk)})_{11} - 2\frac{\hat{c}_{jk,1}\hat{c}_{jk,2}}{(\hat{c}_{jk,1}^2 + \hat{c}_{jk,2}^2)^2} (V_{(jk)})_{12} + \frac{\hat{c}_{jk,1}^2}{(\hat{c}_{jk,1}^2 + \hat{c}_{jk,2}^2)^2} (V_{(jk)})_{22}.
$$

Therefore, the 95 % confidence interval of ϕ_{jk} can be constructed as

$$
\left[\hat{\phi}_{jk} - 1.96\sqrt{v_{(jk)}}, \hat{\phi}_{jk} + 1.96\sqrt{v_{(jk)}} \right].
$$

5 Spectrum

In OSC-DECOMP, each oscillator is described by a VAR(1) model

$$x_t = F x_{t-1} + v_t, \quad v_t \sim \mathrm{N}(0, Q),$$

where

$$F = a \begin{pmatrix} \cos(2\pi f \Delta t) & -\sin(2\pi f \Delta t) \\ \sin(2\pi f \Delta t) & \cos(2\pi f \Delta t) \end{pmatrix}, \quad Q = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}.$$

Note that $F^\top F = F F^\top = a^2 I$. In the following, we summarize results on the spectrum of this model following the notation of [1]. We write $\theta = 2\pi f \Delta t$ for convenience.

The MA(∞) representation of the oscillator model is

$$x_t = \sum_{k=0}^{\infty} F^k v_{t-k}.$$

Thus, the autocovariance of x_t with lag $h \geq 0$ is

$$\Gamma(h) = \mathrm{E}[x_t x_{t-h}^\top] = \sum_{k,l=0}^{\infty} F^k \mathrm{E}[v_{t-k} v_{t-h-l}] (F^\top)^l = \sigma^2 \sum_{k=0}^{\infty} F^{k+h} (F^\top)^k = \sigma^2 F^h \sum_{k=0}^{\infty} a^{2k} = \frac{\sigma^2}{1-a^2} F^h,$$

where we used $F F^\top = a^2 I$. From $\Gamma(-h) = \Gamma(h)^\top$,

$$\Gamma(h) = \frac{\sigma^2}{1-a^2} a^{|h|} \begin{pmatrix} \cos(h\theta) & -\sin(h\theta) \\ \sin(h\theta) & \cos(h\theta) \end{pmatrix}$$

for general h. Then, the spectral density matrix of x_t is

$$f(\lambda) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma(h) \exp(-ih\lambda) = \frac{1}{2\pi} \frac{\sigma^2}{1-a^2} \sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda) \begin{pmatrix} \cos(h\theta) & -\sin(h\theta) \\ \sin(h\theta) & \cos(h\theta) \end{pmatrix}.$$

for $-\pi < \lambda \leq \pi$. By using

$$\begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{pmatrix} \begin{pmatrix} \exp(i\phi) & 0 \\ 0 & \exp(-i\phi) \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ 1/\sqrt{2} & -i/\sqrt{2} \end{pmatrix},$$

we have

$$\begin{aligned} &\sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda) \begin{pmatrix} \cos(h\theta) & -\sin(h\theta) \\ \sin(h\theta) & \cos(h\theta) \end{pmatrix} \\ =& \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{pmatrix} \sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda) \begin{pmatrix} \exp(ih\theta) & 0 \\ 0 & \exp(-ih\theta) \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ 1/\sqrt{2} & -i/\sqrt{2} \end{pmatrix}. \end{aligned}$$

Then, from

$$\begin{aligned} \sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda) \exp(ih\theta) &= \sum_{h=0}^{\infty} a^h \exp(-ih\lambda) \exp(ih\theta) + \sum_{h=0}^{\infty} a^h \exp(ih\lambda) \exp(-ih\theta) - 1 \\ &= \frac{1}{1 - a\exp(i(\theta-\lambda))} + \frac{1}{1 - a\exp(-i(\theta-\lambda))} - 1 \\ &= \frac{1-a^2}{|1 - a\exp(i(\theta-\lambda))|^2}, \end{aligned}$$

$$\sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda)\exp(-ih\theta) = \frac{1-a^2}{|1-a\exp(i(\theta+\lambda))|^2},$$

we obtain

$$\begin{aligned}\sum_{h=-\infty}^{\infty} a^{|h|} \exp(-ih\lambda) \begin{pmatrix} \cos(h\theta) & -\sin(h\theta) \\ \sin(h\theta) & \cos(h\theta) \end{pmatrix} &= (1-a^2) \begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -i/\sqrt{2} & i/\sqrt{2} \end{pmatrix} \begin{pmatrix} \xi & 0 \\ 0 & \eta \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & i/\sqrt{2} \\ 1/\sqrt{2} & -i/\sqrt{2} \end{pmatrix} \\ &= \frac{1-a^2}{2} \begin{pmatrix} \xi+\eta & i(\xi-\eta) \\ -i(\xi-\eta) & \xi+\eta \end{pmatrix},\end{aligned}$$

where $\xi = |1-a\exp(i(\theta-\lambda))|^{-2}$ and $\eta = |1-a\exp(i(\theta+\lambda))|^{-2}$. Hence,

$$f(\lambda) = \frac{\sigma^2}{4\pi} \begin{pmatrix} \xi+\eta & i(\xi-\eta) \\ -i(\xi-\eta) & \xi+\eta \end{pmatrix}. \tag{14}$$

The spectral densities of $x_{t,1}$ and $x_{t,2}$ are

$$f_{11}(\lambda) = f_{22}(\lambda) = \frac{\sigma^2}{4\pi}(\xi+\eta) = \frac{\sigma^2}{2\pi} \frac{1+a^2-2a(\cos\theta)(\cos\lambda)}{|1-2a(\cos\theta)\exp(-i\lambda)+a^2\exp(-2i\lambda)|^{-2}},$$

which is equal to (3.9) in [4]. The cross spectral density of $x_{t,1}$ and $x_{t,2}$ is

$$f_{12}(\lambda) = \frac{\sigma^2}{4\pi} i(\xi-\eta) = \frac{\sigma^2}{2\pi} \frac{2ia(\sin\theta)(\sin\lambda)}{|1-2a(\cos\theta)\exp(-i\lambda)+a^2\exp(-2i\lambda)|^{-2}},$$

which is purely imaginary. Thus, the phase spectrum of $x_{t,1}$ and $x_{t,2}$ is

$$\phi_{12}(\lambda) = \begin{cases} \frac{\pi}{2} & (\sin\theta\sin\lambda > 0) \\ -\frac{\pi}{2} & (\sin\theta\sin\lambda < 0) \end{cases},$$

which indicates that $x_{t,1}$ and $x_{t,2}$ have the opposite phase. The coherency at frequency λ is

$$K_{12}(\lambda) = \frac{f_{12}(\lambda)}{(f_{11}(\lambda)f_{22}(\lambda))^{1/2}} = \frac{2ia(\sin\theta)(\sin\lambda)}{1+a^2-2a(\cos\theta)(\cos\lambda)}.$$

Figure 1 plots the squared coherency $|K_{12}(\lambda)|^2$ for $a = 0.9$ and $\theta = \pi/4$. It indicates that the coherency has a peak at $\lambda = \theta$ and $|K_{12}(\theta)|^2$ is close to one.

6 Whittle likelihood

From the previous section, the spectral density matrix of the observation $y_t = Hx_t + w_t$ with $w_t \sim \mathrm{N}(0, R)$ is

$$f_Y(\lambda) = \sum_{k=1}^{K} H_k f_X^{(k)}(\lambda) H_k^\top + \frac{1}{2\pi} R,$$

where $H = (H_1, H_2, \ldots, H_K)$ and $f_X^{(k)}(\lambda)$ is the spectral density matrix of the k-th oscillator (14). Then, the multivariate Whittle log-likelihood [9] for $y_1, \ldots, y_T$ is

$$\log W(\theta) = -\sum_{k=1}^{T/2} (\log\det(2\pi f_Y(\omega_k)) + Y_k^* (2\pi f_Y(\omega_k))^{-1} Y_k)$$

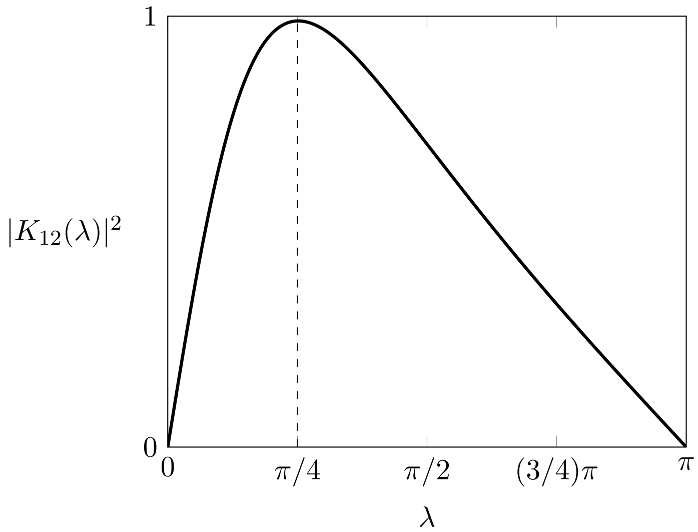

Figure 1: Squared coherency $|K_{12}(\lambda)|^2$

where $\omega_k = 2\pi k/T$ and $Y_1, \ldots, Y_{T/2}$ is the periodogram defined by

$$Y_k = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} y_t \exp(-it\omega_k).$$

It is used for setting the initial guess of $\sigma_1^2, \ldots, \sigma_K^2$ and τ^2.

References

[1] Brockwell, P. J. & Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer.

[2] Efron, B. & Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65**, 457–483.

[3] Kitagawa, G. (2020). Computation of the gradient and Hessian of the log-likelihood of the state-space model by the Kalman filter. arXiv:2011.09638.

[4] Matsuda, T. & Komaki, F. (2017). Time series decomposition into oscillation components and phase estimation. *Neural Computation* **29**, 332–367.

[5] Matsuda, T. & Komaki, F. (2017). Multivariate time series decomposition into oscillation components. *Neural Computation* **29**, 2055–2075.

[6] Matsuda, T., Homae, F., Watanabe, H., Taga, G. & Komaki, F. (2022). Oscillator decomposition of infant fNIRS data. *PLOS Computational Biology*, to appear.

[7] Rasmussen, C. E. (2006). Conjugate gradient algorithm. Matlab code version 2006-09-08. `http://learning.eng.cam.ac.uk/carl/code/minimize/minimize.m`

[8] Shumway, R. H. & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis* **3**, 253–264.

[9] WHITTLE, P. (1953) Estimation and information in stationary time series. *Arkiv for Matematik* **2**, 423–434.