Supplementary Material for EMORF-II

This document contains the supplementary material for the manuscript titled "**EMORF-II: ADAPTIVE EM-BASED OUTLIER-ROBUST FILTERING WITH CORRELATED MEASUREMENT NOISE**" submitted to the *"International Workshop on Machine Learning and Signal Processing 2025"*. It provides notations used in the paper, details of derivations of $\hat{\mathcal{I}}_k^i$ and $\hat{b}_k$ in the paper and discussion of some more numerical evaluations.

Notation

As a general notation in this work, $\mathbf{r}^\top$ is the transpose of the vector $\mathbf{r}$, r^i denotes the i-th element of a vector $\mathbf{r}$; $\mathbf{r}^{-i}$ is the vector $\mathbf{r}$ with its i-th element removed; the subscript k is used for the time index; $\mathbf{r}_k$ is the vector $\mathbf{r}$ at time instant k; $\mathbf{r}_k^-$ is the sequence of vectors $\mathbf{r}$ over all time steps except k; $R^{i,j}$ is the element of the matrix $\mathbf{R}$ in the i-th row and j-th column; $\mathbf{R}^{-1}$ is the inverse of $\mathbf{R}$; $|\mathbf{R}|$ is the determinant of $\mathbf{R}$; $\delta(\cdot)$ represents the delta function; $\langle\cdot\rangle_{q(\psi_k)}$ denotes expectation with respect to the distribution $q(\psi_k)$; $\mathrm{tr}(\cdot)$ is the trace operator; $a \bmod b$ denotes the remainder of a/b; the superscripts $^-$ and $^+$ mark predicted and updated filtering parameters, respectively. Furthermore, for any matrix $\mathbf{R}$ we obtain the sub-matrix $\mathbf{R}^{-i,-i}$ by removing its i-th column and row. $\mathbf{R}_k(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{i-})$ denotes $\mathbf{R}_k(\mathcal{I}_k)$ evaluated at $\mathcal{I}_k$ with its i-th element as $\mathcal{I}_k^i$ and remaining entries $\hat{\boldsymbol{\mathcal{I}}}_k^{i-}$.

Details of estimating $\hat{\mathcal{I}}_k^i$

Invoking the EM algorithm, the estimates $\hat{\mathcal{I}}_k^i$ can be obtained with the following maximization

$$\hat{\mathcal{I}}_k^i = \underset{\mathcal{I}_k^i}{\mathrm{argmax}} \Big\langle \ln(p(\mathrm{x}_k, \mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{i-}, \hat{b}_k | \mathbf{y}_{1:k}) \Big\rangle_{q(\mathbf{x}_k)}$$

This leads to Equation (21) in the manuscript given as

$$\hat{\mathcal{I}}_k^i = \underset{\mathcal{I}_k^i}{\mathrm{argmax}} \Big\{ \underbrace{-\tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\, \mathbf{R}_k^{-1}(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{i-})\big) - \tfrac{1}{2}\,\ln\big|\mathbf{R}_k(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{i-})\big| + \ln\big((1-\theta_k)\, f(a_k, \hat{b}_k)\, (\mathcal{I}_k^i)^{a_k-1}\, e^{-\hat{b}_k \mathcal{I}_k^i} + \theta_k\, \delta(\mathcal{I}_k^i - 1)\big) + k_1^i}_{\ln q(\mathcal{I}_k^i)} \Big\} \tag{21}$$

where the term to be maximized is supposed as $\ln q(\mathcal{I}_k^i)$. We perform the following manipulations to obtain Equations (23)-(25) in the manuscript. Manipulating $\ln q(\mathcal{I}_k^i)$ we can get

$$\ln q(\mathcal{I}_k^i) = -\tfrac{1}{2}\ln\big|\mathbf{R}_k(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big| - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\, \mathbf{R}_k^{-1}(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big) + \ln\big((1-\theta_k^i)\, f(a_k, \hat{b}_k)\, (\mathcal{I}_k^i)^{a_k-1}\, e^{-\hat{b}_k\, \mathcal{I}_k^i} \;+\; \theta_k^i\, \delta(\mathcal{I}_k^i - 1)\big) + k_1^i$$

which leads to

$$\begin{aligned} q(\mathcal{I}_k^i) &= k_2^i \exp\big(-\tfrac{1}{2}\ln\big|\mathbf{R}_k(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big| - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\, \mathbf{R}_k^{-1}(\mathcal{I}_k^i, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big)\big) \times (1-\theta_k^i)\, f(a_k, \hat{b}_k)\, (\mathcal{I}_k^i)^{a_k-1}\, e^{-\hat{b}_k\, \mathcal{I}_k^i} \\ &\quad + k_2^i \exp\big(-\tfrac{1}{2}\ln\big|\mathbf{R}_k(\mathcal{I}_k^i = 1, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big| - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\, \mathbf{R}_k^{-1}(\mathcal{I}_k^i = 1, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big)\big)\, \theta_k^i\, \delta(\mathcal{I}_k^i - 1) \end{aligned}$$

where k_2^i is another proportionality constant. Further manipulation of $q(\mathcal{I}_k^i)$ yields

$$\begin{aligned} q(\mathcal{I}_k^i) &= k_2^i (R_k^{ii}/\mathcal{I}_k^i)^{-\frac{1}{2}}\, \big|\hat{\mathbf{R}}_k^{-i,-i}\big|^{-\frac{1}{2}} \exp\big(-\tfrac{1}{2}\, W_k^{ii}\, \tfrac{\mathcal{I}_k^i}{R_k^{ii}} - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k^{-i,-i}\, (\hat{\mathbf{R}}_k^{-i,-i})^{-1}\big)\big) \times (1-\theta_k^i)\, f(a_k, \hat{b}_k)\, (\mathcal{I}_k^i)^{a_k-1}\, e^{-\hat{b}_k \mathcal{I}_k^i} \\ &\quad + k_2^i \exp\big(-\tfrac{1}{2}\ln\big|\mathbf{R}_k(\mathcal{I}_k^i = 1, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big| - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\, \mathbf{R}_k^{-1}(\mathcal{I}_k^i = 1, \hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big)\big)\, \theta_k^i\, \delta(\mathcal{I}_k^i - 1) \end{aligned}$$

Moving the relevant terms in the exponents we get

$$q\big(\mathcal{I}_k^i\big) = k_2^i (R_k^{ii})^{-\frac{1}{2}} \left|\hat{\mathbf{R}}_k^{-i,-i}\right|^{-\frac{1}{2}} \exp\!\big(-\tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k^{-i,-i}\,(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\big)\big) \times (1-\theta_k^i)\, f(a_k,\hat{b}_k)\,(\mathcal{I}_k^i)^{\overbrace{\left(a_k+\frac{1}{2}\right)}^{\alpha_k}-1}\, e^{-\overbrace{\left(\hat{b}_k+\frac{1}{2}\frac{W_k^{ii}}{R_k^{ii}}\right)}^{\beta_k^i}\mathcal{I}_k^i}$$
$$+\, k_2^i \exp\!\big(-\tfrac{1}{2}\ln\!\big|\mathbf{R}_k(\mathcal{I}_k^i=1,\hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big| - \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\,\mathbf{R}_k^{-1}(\mathcal{I}_k^i=1,\hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big)\big)\,\theta_k^i\,\delta(\mathcal{I}_k^i-1)$$

α_k and β_k^i can easily be identified as follows given as Equations (26)-(27) in the manuscript

$$\alpha_k = a_k + 0.5 \tag{26}$$

$$\beta_k^i = \hat{b}_k + 0.5\frac{W_k^{ii}}{R_k^{ii}} \tag{27}$$

Through further manipulation we can write $q(\mathcal{I}_k^i)$ as

$$q(\mathcal{I}_k^i) = k_2^i\ \underbrace{\big(R_k^{ii}\big)^{-\frac{1}{2}}\big|\hat{\mathbf{R}}_k^{-i,-i}\big|^{-\frac{1}{2}} \exp\!\big(-\tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k^{-i,-i}(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\big)\big)(1-\theta_k^i)\,\frac{\Gamma(\alpha_k)\hat{b_k}^{a_k}}{\Gamma(a_k)\beta_k^{i\,\alpha_k}}}_{G_k^i}\ \overbrace{f(\alpha_k,\beta_k^i)\,(\mathcal{I}_k^i)^{\alpha_k-1}\,e^{-\beta_k^i\,\mathcal{I}_k^i}}^{Gamma(\alpha_k,\beta_k^i)}$$
$$+\ k_2^i\ \underbrace{\exp\!\big(-\tfrac{1}{2}\ \ln\!\big|\mathbf{R}_k(\mathcal{I}_k^i=1,\hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big|\ -\ \tfrac{1}{2}\,\mathrm{tr}\big(\mathbf{W}_k\,\mathbf{R}_k^{-1}(\mathcal{I}_k^i=1,\hat{\boldsymbol{\mathcal{I}}}_k^{-i})\big)\big)\,\theta_k^i}_{H_k^i}\ \delta(\mathcal{I}_k^i-1)$$

Since maximizing logarithm of a positive function is equivalent to maximizing the function, the resulting decision criterion (from Equation 21) depends on which part - G_k^i or H_k^i - of the multi-modal function $q(\mathcal{I}_k^i)$ has higher weight. If $H_k^i/G_k^i \geq 1$ we declare $\hat{\mathcal{I}}_k^i = 1$. Otherwise, if $H_k^i/G_k^i < 1$, we assign $\hat{\mathcal{I}}_k^i$ as the the mode of $Gamma(\alpha_k, \beta_k^i)$. The criterion is summarized in Equation (23) in the manuscript reproduced as

$$\mathcal{I}_k^i = \begin{cases} 1 & \text{if } H_k^i/G_k^i \geq 1 \\ (\alpha_k - 1)/\beta_k^i & \text{if } H_k^i/G_k^i < 1 \end{cases} \tag{23}$$

Details of estimating $\hat{b}_k$

Employing the EM algorithm the estimate for b_k can be calculated through the following maximization

$$\hat{b}_k = \underset{b_k^i}{\operatorname{argmax}} \big\langle \ln(p(\mathrm{x}_k, \hat{\boldsymbol{\mathcal{I}}}_k, b_k|\mathbf{y}_{1:k})\big\rangle_{q(\mathbf{x}_k)}$$

which leads to

$$\hat{b}_k = \underset{b_k}{\operatorname{argmax}} \sum_{i=1}^{m} \ln\Big[(1-\theta_k^i) f(a_k, b_k)(\hat{\mathcal{I}}_k^i)^{a_k-1} e^{-b_k\hat{\mathcal{I}}_k^i}\ + \theta_k^i \delta(\hat{\mathcal{I}}_k^i - 1)\Big] + \ln(f(A_k, B_k) b_k^{A_k-1} e^{-B_k b_k})$$

Since $\hat{\mathcal{I}}_k^i = 1$ do not affect the estimates $\hat{b}_k$, we can further write

$$\hat{b}_k = \underset{b_k}{\operatorname{argmax}} \sum_{\{i:\hat{\mathcal{I}}_k^i \neq 1\}} \ln\Big[(1-\theta_k^i) f(a_k, b_k)(\hat{\mathcal{I}}_k^i)^{a_k-1} e^{-b_k\hat{\mathcal{I}}_k^i}\Big] + \ln(f(A_k, B_k) b_k^{A_k-1} e^{-B_k b_k})$$

which can be simplified as

$$\hat{b}_k = \underset{b_k}{\operatorname{argmax}} \sum_{\{i:\hat{\mathcal{I}}_k^i \neq 1\}} (a_k \ln b_k - b_k \mathcal{I}_k^i) - B_k b_k + (A_k - 1)\ln b_k$$

Maximizing the above equation through differentiation leads to the estimate $\hat{b}_k$ given as

$$\hat{b}_k = \frac{M_k a_k + A_k - 1}{B_k + \sum_{\{i:\hat{\mathcal{I}}_k^i \neq 1\}} \hat{\mathcal{I}}_k^i}$$

where $M_k = \#\{\, i : \hat{\mathcal{I}}_k^i \neq 1\}$ i.e. the count of $\boldsymbol{\mathcal{I}_k}$ elements not equal to one. Further defining $\overline{A}_k = M_k a_k + A_k$ and $\overline{B_k} = B_k + \sum_{\{i:\mathcal{I}_k^i \neq 1\}} \hat{\mathcal{I}}_k^i$ results in Equation (28) of the manuscript

$$\hat{b}_k = \frac{\overline{A}_k - 1}{\overline{B}_k} \tag{28}$$

Other numerical experiments

We further extend our numerical experiments by introducing outliers generated from Uniform and Laplace distributions, replacing the previously considered Gaussian-based outliers. Across a range of parameter variations, the comparative performance trends remained similar to those presented in the main manuscript. Further evaluation of EMORF-II was conducted by testing alternative prior parameters θ_k^i, including values sampled from a Uniform distribution over the range $[0.05, 0.95]$. The results show that EMORF-II maintains robust performance across variations in θ_k^i. Nevertheless, setting $\theta_k^i = 0.5$ is recommended as a default choice in the absence of specific prior knowledge regarding outlier rates.