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"*. This document goes over the derivation for results present in the paper with more detail to assist any interested reader in understanding how the update equations in the paper are formulated. This document can be found in the repository here.

0.1. Derivation of $q(\mathbf{x}_k)$

In deriving the distribution $q(\mathbf{x}_k)$, we discard from its VB approximation all terms that do not depend on $\mathbf{x}_k$.

$$q(\mathbf{x}_k) \propto \exp\Bigg(-\frac{1}{2}(\mathbf{y}_k - h(\mathbf{x}_k))^\top \mathbf{R}_k^{-1}(\hat{\mathcal{I}}_k)(\mathbf{y}_k - h(\mathbf{x}_k)) - \frac{1}{2}(\mathbf{x}_k - \mathbf{m}_k^-)^\top (\mathbf{P}_k^-)^{-1}(\mathbf{x}_k - \mathbf{m}_k^-)\Bigg) \tag{1}$$

where $\mathbf{R}_k^{-1}(\hat{\mathcal{I}}_k)$ is the matrix obtained by inverting $\mathbf{R}(\mathcal{I}_k)$ after plugging in the estimate of $\mathcal{I}_k$. Finally to approximate $q(\mathbf{x}_k)$ as the normal distribution $\mathcal{N}(\mathbf{x}_k|\mathbf{m}_k^+, \mathbf{P}_k^+)$ with mean $\mathbf{m}_k^+$ and covariance $\mathbf{P}_k^+$ we employ then the general Gaussian filtering equations which provide

$$\mathbf{m}_k^+ = \mathbf{m}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mu_k) \tag{2}$$
$$\mathbf{P}_k^+ = \mathbf{P}_k^- - \mathbf{C}_k\mathbf{K}_k^\top \tag{3}$$

where

$$\mathbf{K}_k = \mathbf{C}_k(\mathbf{U}_k + \mathbf{R}_k(\hat{\mathcal{I}}_k))^{-1} \tag{4}$$
$$\mu_k = \int h(\mathbf{x}_k)\mathcal{N}(\mathbf{x}_k|\mathbf{m}_k^-, \mathbf{P}_k^-)\, d\mathbf{x}_k \tag{5}$$
$$\mathbf{U}_k = \int (h(\mathbf{x}_k) - \mu_k)(h(\mathbf{x}_k) - \mu_k)^\top \mathcal{N}(\mathbf{x}_k|\mathbf{m}_k^-, \mathbf{P}_k^-)\, d\mathbf{x}_k \tag{6}$$
$$\mathbf{C}_k = \int (\mathbf{x}_k - \mathbf{m}_k^-)(h(\mathbf{x}_k) - \mu_k)^\top \mathcal{N}(\mathbf{m}_k|\mathbf{m}_k^-, \mathbf{P}_k^-)\, d\mathbf{x}_k \tag{7}$$

0.2. Derivation for $q(\mathcal{I}_k^i)$

We estimate $q(\mathcal{I}_k^i)$ by employing the Expectation Maximization (EM) Algorithm as

$$\hat{\mathcal{I}}_k^i = \underset{\mathcal{I}_k^i}{\operatorname{argmax}} \langle \ln(p(\mathbf{x}_k, \mathcal{I}_k^i, \hat{\mathcal{I}}_k^{i-}|\mathbf{y}_{1:k}))\rangle_{q(\mathbf{x}_k)q(b_k)} \tag{8}$$

By excluding all the elements that are not relevant to $\mathcal{I}_k^i$ we get the following equation

$$\hat{\mathcal{I}}_k^i = \underset{\mathcal{I}_k^i}{\operatorname{argmax}} \Big\{ -\frac{1}{2}\operatorname{tr}\big(\mathbf{W}_k\mathbf{R}_k^{-1}(\mathcal{I}_k^i, \hat{\mathcal{I}}_k^{i-})\big) - \frac{1}{2}\ln|\mathbf{R}_k(\mathcal{I}_k^i, \hat{\mathcal{I}}_k^{i-})| + \ln\big((1-\theta_k)f(a_k, \hat{b}_k)(\mathcal{I}_k^i)^{a_k - 1}e^{-\hat{b}_k\mathcal{I}_k^i} + \theta_k\delta(\mathcal{I}_k^i - 1)\big)\Big\} \tag{9}$$

$\mathbf{R}_k(\mathcal{I}_k^i, \hat{\mathcal{I}}_k^{i-})$ 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 $\hat{\mathcal{I}}_k^{i-}$ and

$$\mathbf{W}_k = \int (\mathbf{y}_k - \mathbf{h}(\mathbf{x}_k))(\mathbf{y}_k - \mathbf{h}(\mathbf{x}_k))^\top \mathcal{N}(\mathbf{x}_k|\mathbf{m}_k^+, \mathbf{P}_k^+) d\mathbf{x}_k \tag{10}$$

We perform the following evaluations to estimate which value of $\mathcal{I}_k^i$ maximizes our objective function. Let

$$f_1(\mathcal{I}_k^i) = -\frac{1}{2}\ln\left(\left|\mathbf{R}_k(\mathcal{I}_k^i, \hat{\mathcal{I}}_k^{-i})\right|\right),$$
$$f_2(\mathcal{I}_k^i) = -\frac{1}{2}\operatorname{tr}\left(\mathbf{W}_k\mathbf{R}_k^{-1}(\mathcal{I}_k^i, \hat{\mathcal{I}}_k^{-i})\right),$$
$$f_3(\mathcal{I}_k^i) = f_1(\mathcal{I}_k^i) + f_2(\mathcal{I}_k^i), \quad \ln(f_4(\mathcal{I}_k^i)) = f_3(\mathcal{I}_k^i)$$
$$\ln(q(\mathcal{I}_k^i)) \propto \ln(f_4(\mathcal{I}_k^i)) + \ln\left((1-\theta_i)f(a_k, \hat{b}_k){\mathcal{I}_k^i}^{(a_k-1)}e^{-\hat{b}_k\mathcal{I}_k^i}\right) + \theta_i\delta(\mathcal{I}_k^i - 1) \tag{11}$$
$$q(\mathcal{I}_k^i) \propto f_4(\mathcal{I}_k^i \neq 1)(1-\theta_i)a_k, \hat{b}_k){\mathcal{I}_k^i}^{(a_k-1)}e^{-\hat{b}_k\mathcal{I}_k^i} + f_4(\mathcal{I}_k^i = 1)\theta_i\delta(\mathcal{I}_k^i - 1). \tag{12}$$

$$f_4(\mathcal{I}_k^i \neq 1) = \left(\frac{R_k^{ii}}{\mathcal{I}_k^i}\right)^{-\frac{1}{2}} |\hat{\mathbf{R}}_k^{-i,-i}|^{-\frac{1}{2}} e^{\left(-\frac{1}{2}W_k^{ii}\left(\frac{\mathcal{I}_k^i}{\mathbf{R}_k^{ii}}\right) - \frac{1}{2}\operatorname{tr}\left(\mathbf{W}_k^{-i,-i}(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\right)\right)} \tag{13}$$

$$q(\mathcal{I}_k^i) \propto (R_k^{ii})^{-\frac{1}{2}}|\hat{\mathbf{R}}_k^{-i,-i}|^{-\frac{1}{2}}e^{-\frac{1}{2}\operatorname{tr}\left(\mathbf{W}_k^{-i,-i}(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\right)} (1-\theta_i)f(a_k, \hat{b}_k){\mathcal{I}_k^i}^{a_k - 1 + 0.5}e^{-\mathcal{I}_k^i(\hat{b}_k + 0.5\frac{W_k^{ii}}{R_k^{ii}})} + f_4(\mathcal{I}_k^i = 1)\theta_i\delta(\mathcal{I}_k^i - 1) \tag{14}$$

$$\alpha_k = a_k + 0.5; \quad \beta_i = \hat{b}_k + 0.5\frac{W_k^{ii}}{R_k^{ii}}$$

Here $\mathbf{W}_k^{-i,-i}$ and $\hat{\mathbf{R}}_k^{-i,-i}$ can be evaluated from $\mathbf{W}_k$ and $\mathbf{R}_k(\mathcal{I}_k)$ by removing the i-th row and column. Furthermore, α and $\beta_i \quad \forall i$ represent updated parameters for our posterior distribution for $\mathcal{I}_k$

$$q(\mathcal{I}_k^i) = k_i(R_k^{ii})^{-\frac{1}{2}}|\hat{\mathbf{R}}_k^{-i,-i}|^{-\frac{1}{2}}e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{W}_k^{-i,-i}(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\right)}$$
$$(1-\theta_i)a_k,\hat{b})\mathcal{I}_k^{i^{\alpha_k-1}}e^{-\mathcal{I}_k^i\beta_i} + k_i f_4(\mathcal{I}_k^i = 1)\theta_i\delta(\mathcal{I}_k^i - 1) \tag{15}$$

$$q\left(\mathcal{I}_k^i\right) = \left(1-\Omega_k^i\right) f\left(\alpha,\beta_i\right)\left(\mathcal{I}_k^i\right)^{\alpha-1} e^{-\beta_i\mathcal{I}_k^i} + \Omega_k^i\delta\left(\mathcal{I}_k^i - 1\right) \tag{16}$$

Where $\Omega_k^i = k_i f_4(\mathcal{I}_k^i = 1)\theta_i$ (17)

By exploiting the fact that the distribution in (15) must add to 1 we can calculate k_i as

$$k_i = \frac{1}{H(\alpha,\beta_i)} \tag{18}$$

$$H(\alpha,\beta_i) = (R_k^{ii})^{-\frac{1}{2}}|\hat{\mathbf{R}}_k^{-i,-i}|^{-\frac{1}{2}}e^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{W}_k^{-i,-i}(\hat{\mathbf{R}}_k^{-i,-i})^{-1}\right)}$$
$$(1-\theta_i)f(a_k,\hat{b}_k)\mathcal{I}_k^{i^{\alpha_k-1}}e^{-\mathcal{I}_k^i\beta_i} + f_4(\mathcal{I}_k^i = 1)\theta_i \tag{19}$$

Thereby, using (18) and (17) we can estimate Ω_k^i to be

$$\Omega_k^i = \frac{f_4(\mathcal{I}_k^i = 1)\theta_i}{H(\alpha,\beta_i)} \tag{20}$$

We shall employ the following criterion to estimate a value for $\mathcal{I}_k^i$

$$\forall i\, \mathcal{I}_k^i = \begin{cases} 1 & \text{if } \Omega_k^i \geq 0.5 \\ \frac{\alpha-1}{\beta_i} & \text{if } \Omega_k^i < 0.5 \end{cases} \tag{21}$$

0.3. Derivation of $q(b_k)$

Similar to the case for $q(\mathcal{I}_k^i)$, we shall assume a point distribution on b_k i.e $q(b_k) = \delta(b_k - \hat{b}_k)$, which will reduce the estimation of b_k to the following M-Step of the EM algorithm.

$$b_k = \operatorname*{argmax}_{b_k} \langle \ln p(\mathbf{x}_k, \mathcal{I}_k, b_k|\mathbf{y}_k)\rangle_{q(\mathbf{x}_k)q(\mathcal{I}_k)} \tag{22}$$

$$b_k = \operatorname*{argmax}_{b_k}\langle \ln\left(p(\mathbf{y}_k|\mathcal{I}_k x_k)\right)\ln\left(p(x)\right) + \ln\left(p(\mathcal{I}_k|b_k)\right) +$$
$$\ln\left(p(b_k)\right)\rangle_{q(\mathcal{I}_k)q(\mathbf{x}_k)} \tag{23}$$

Removing parts of the equation that do not depend on b_k

$$b_k = \operatorname*{argmax}_{b_k}\sum_{i=1}^{m}\ln\Big[(1-\theta_i)a_k,b_k)(\hat{\mathcal{I}}_k^i)^{a_k-1}e^{-b_k\hat{\mathcal{I}}_k^i}$$
$$+\theta_i\delta(\hat{\mathcal{I}}_k^i - 1)\Big] + \ln(f(A_k,B_k))b_k^{A-1}e^{-B_k b_k} \tag{24}$$

Since we are only interested in the case where $\hat{\mathcal{I}}_k^i \neq 1$ our equation reduces to

$$b_k = \operatorname*{argmax}_{b_k}\sum_{i=1}^{m}\ln\Big[(1-\theta_i)a_k,b_k)(\hat{\mathcal{I}}_k^i)^{a_k-1}e^{-b_k\hat{\mathcal{I}}_k^i}\Big]$$
$$+\ln(f(A_k,B_k))b_k^{A-1}e^{-B_k b_k} \tag{25}$$

$$\hat{b}k = \frac{Ma + A - 1}{B + \sum i \in i : \mathcal{I}_k^i \neq 1\mathcal{I}_k^i} \tag{26}$$

$$\hat{b}_k = \frac{\overline{A} - 1}{\overline{B}} \tag{27}$$

We obtain our estimate for b_k by maximizing the likelihood, which yields

$$\hat{b}_k = \frac{Ma + A - 1}{B + \sum_{i\in\{i:\mathcal{I}_k^i\neq 1\}}\mathcal{I}_k^i}. \tag{28}$$

For clarity, we define

$$\overline{A} = Ma + A \quad \text{and} \quad \overline{B} = B + \sum_{i\in\{i:\mathcal{I}_k^i\neq 1\}}\mathcal{I}_k^i.$$

Thus, the estimator can be rewritten as

$$\hat{b}_k = \frac{\overline{A} - 1}{\overline{B}}. \tag{29}$$

In this formulation, the summation in $\overline{B}$ includes only those indices i for which $\mathcal{I}_k^i \neq 1$, thereby excluding any contributions from observations where $\mathcal{I}_k^i = 1$.