$$
\begin{gathered}
\mathbb{E}_{q(\Xi)}[\log \mathsf{P}(X,S)] = \\
\mathbb{E}_{q(\Xi)}[\log \mathsf{P}(X \mid S) + \log \mathsf{P}(S)] = \\
\Big[\sum_{t=1}^{T} \mathbb{E}_{q(\Xi_t)} \log[\mathsf{P}(X_t \mid S_t)] + \sum_{t=1}^{T} \mathbb{E}_{q(\Xi_t)}[\log \mathsf{P}(S_t)]\Big] = \\
-\mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XX^*]\right) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\mathrm{A}^*(\theta)\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XS^*]\right) \\
+ \mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big) + \log|\mathbf{\Gamma}| + \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{q(\Xi)}[SS^*]\big)\Big].
\end{gathered}
$$

M-шаг для неизвестного шума

Пусть $\mathbf{\Lambda} = h\mathrm{I}_{\mathrm{L}}$, и требуется найти h, максимизирующий следующую функцию:

$$
\begin{gathered}
-\mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XX^*]\right) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\mathrm{A}^*(\theta)\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XS^*]\right) \\
+ \mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big) + \log|\mathbf{\Gamma}| + \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{q(\Xi)}[SS^*]\big)\Big].
\end{gathered}
$$

Исключим слагаемые, не содержащие $\mathbf{\Lambda}$:

$$
\begin{gathered}
-\mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XX^*]\right) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\mathrm{A}^*(\theta)\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XS^*]\right) \\
+ \mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big)\Big].
\end{gathered}
$$

При этом, $W^{-1} = \frac{1}{h}\mathrm{I}_{\mathrm{L}}$. Проведем преобразования:

$$
\begin{gathered}
-\mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XX^*]\right) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\mathrm{A}^*(\theta)\mathbf{\Lambda}^{-1}\mathbb{E}_{q(\Xi)}[XS^*]\right) \\
+ \mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big)\Big] = \\
-\mathrm{T}\left[\log|h\mathrm{I}_{\mathrm{L}}| + \mathrm{Tr}\left(\frac{1}{h}\mathrm{I}_{\mathrm{L}}\mathbb{E}_{q(\Xi)}[XX^*]\right) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\frac{1}{h}\mathrm{I}_{\mathrm{L}}\mathrm{A}^*(\theta)\mathbb{E}_{q(\Xi)}[XS^*]\right)\right. \\
\left. + \mathrm{Tr}\left(\frac{1}{h}\mathrm{I}_{\mathrm{L}}\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\right)\right] = \\
-\mathrm{T}\left[\mathrm{L}\log h + \frac{1}{h}\Big(\mathrm{Tr}\big(\mathbb{E}_{q(\Xi)}[XX^*]\big) - 2\,\mathrm{Re}\,\mathrm{Tr}\big(\mathrm{A}^*(\theta)\mathbb{E}_{q(\Xi)}[XS^*]\big)\right. \\
\left. + \mathrm{Tr}\big(\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big)\Big)\right].
\end{gathered}
$$

Обозначим через B выражение

$$
\mathrm{Tr}\big(\mathbb{E}_{q(\Xi)}[XX^*]\big) - 2\,\mathrm{Re}\,\mathrm{Tr}\left(\mathrm{A}^*(\theta)\mathbb{E}_{q(\Xi)}[XS^*]\right) + \mathrm{Tr}\big(\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big).
$$

Получаем функцию

$$
\mathcal{S}(h) = -\mathrm{T}\left[\mathrm{L}\log h + \frac{1}{h}B\right].
$$

Производная этой формулы имеет следующий вид:

$$
\mathcal{S}'(h) = \frac{B}{h^2} - \frac{\mathrm{L}}{h}.
$$

Эта производная равна нулю, если:

$$
h = \frac{B}{\mathrm{L}}.
$$

Соответственно:

$$h^* = \frac{1}{\mathrm{L}}\Big(\operatorname{Tr}\big(\mathbb{E}_{q(\Xi)}[XX^*]\big) - 2\operatorname{Re}\operatorname{Tr}\big(\mathrm{A}^*(\theta)\mathbb{E}_{q(\Xi)}[XS^*]\big) + \operatorname{Tr}\big(\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big)\Big).$$

Ввиду того, что нет гарантии положительности B, нужно брать максимум из двух чисел: вышеуказанной оценки и некоторого малого числа ε.

$$h^{(\tau)} = \max\left(\frac{1}{\mathrm{L}}\Big(\operatorname{Tr}\big(\mathbb{E}_{q(\Xi)}[XX^*]\big) - 2\operatorname{Re}\operatorname{Tr}\big(\mathrm{A}^*(\theta)\mathbb{E}_{q(\Xi)}[XS^*]\big) + \operatorname{Tr}\big(\mathrm{A}(\theta)\mathbb{E}_{q(\Xi)}[SS^*]\mathrm{A}^*(\theta)\big)\Big), \varepsilon\right).$$

MAP-оценивание ковариации сигналов

Можно превратить EM в MAP-EM, если искать не $\log P(X \mid \theta)$, а $\log P(X \mid \theta) + \log P(\theta)$, в случае моей задачи разумно задать априор для ковариации сигналов (EM-алгоритм с ММП-оценкой ее не ищет практически, неидентифицируемость или слабая идентифицируемость масштаба, вероятно). Будем считать, что ковариация сигналов $\mathbf{\Gamma}_{p\times p}$ имеет обратное комплексное распределение Вишарта:

$$\mathbf{\Gamma} \sim \mathcal{CW}^{-1}(\mathbf{\Psi}, \nu, p), \tag{1}$$

где Ψ – априорная информация о ковариации сигналов, ν – степень уверенности.
В таком случае, оптимизации подвергается следующая функция:

$$\mathcal{Q}(\theta|\theta^{(\tau-1)}) = \mathbb{E}_{q(\Xi)}[\log P(X, S|\theta)] + \log P(S). \tag{2}$$

Точный шаг для обновления параметров будет определяться так:

$$\mathbf{\Gamma} = \frac{\mathrm{T}\mathbb{E}[SS^* \mid q(\Xi)] + \Psi}{\nu + d + \mathrm{T}}. \tag{3}$$

Чем больше ν, тем больше уверенности в априорных данных. Плотность для $\mathbf{\Gamma}$ пропорциональна следующей величине (нормировочная константа в данном случае не важна):

$$p(\mathbf{\Gamma}) \propto |\mathbf{\Gamma}|^{-(\nu+d+1)/2} \exp\big(-\tfrac{1}{2}\operatorname{tr}(\Psi\mathbf{\Gamma}^{-1})\big). \tag{4}$$