EM-алгоритм, стохастическая модель сигнала

§1 Постановка проблемы

Предположим, имеется линейная антенная решетка, состоящая из L сенсоров, которая принимает сигналы, направленные из K источников, причем $\mathrm{K} < \mathrm{L}$. Этим источникам соответствуют угловые координаты (DoA) θ, практически не изменяющиеся во времени. По итогам измерений было получено T снимков полученного сигнала, причем ввиду стохастических технических сбоев, связанных с сенсорами, большая часть таких снимков содержит помимо надежных данных ненадежные, которые в рамках данной задачи рассматриваются как пропуски. Пусть X — полный набор наблюдений (сигналов, полученных сенсорами в моменты времени $t = 1, \dots, \mathrm{T}$, X_t соответствует наблюдению в момент времени t, через x и x_t будем обозначать реализации полного набора наблюдений и наблюдения в отдельный момент времени t соответственно. Ввиду наличия пропусков в данных, будем считать, что X состоит из наблюдаемой части $X_o = \{X_{t,o_t}\}_{t=1}^{\mathrm{T}}$ и ненаблюдаемой: $X_m = \{X_{t,m_t}\}_{t=1}^{\mathrm{T}}$, причем $o_t \cup m_t = \{1, \dots, \mathrm{L}\}, o_t \cap m_t = \varnothing, \forall t \in \{1, \dots, \mathrm{T}\}$. Предполагается, что $\nexists o_t : |o_t| = 0$, т.е. нет таких наблюдений, которые состоят лишь из ненаблюдаемой части.

Набор наблюдений X является результатом следующей модели наблюдений:

$$X = \mathrm{A}S + N, \tag{1}$$

где

$$X = [X_1, \dots, X_{\mathrm{T}}], S = [S_1, \dots, S_{\mathrm{T}}], N = [N_1, \dots, N_{\mathrm{T}}], \tag{2}$$

и $x_t \in \mathbb{C}^{L\times \mathrm{T}}, s_t \in \mathbb{C}^{K\times \mathrm{T}}, n_t \in \mathbb{C}^{L\times \mathrm{T}}$ – векторы-столбцы, соответствующие наблюдениям, источникам и шумам в момент времени $t = 1, \dots, \mathrm{T}$, A – матрица векторов направленности для равномерного линейного массива:

$$\mathrm{A}(\theta) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-2j\pi\frac{\mathrm{d}}{\lambda}\sin(\theta_1)} & e^{-2j\pi\frac{\mathrm{d}}{\lambda}\sin(\theta_2)} & \dots & e^{-2j\pi\frac{\mathrm{d}}{\lambda}\sin(\theta_{\mathrm{K}})} \\ \dots & \dots & \ddots & \vdots \\ e^{-2j\pi(\mathrm{L}-1)\frac{\mathrm{d}}{\lambda}\sin(\theta_1)} & e^{-2j\pi(\mathrm{L}-1)\frac{\mathrm{d}}{\lambda}\sin(\theta_2)} & \dots & e^{-2j\pi(\mathrm{L}-1)\frac{\mathrm{d}}{\lambda}\sin(\theta_{\mathrm{K}})} \end{bmatrix}.$$

Сигналы, испускаемые источниками, шумы на сенсорах и наблюдения предполагаются стохастическими: $S_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{K}\times 1}, \mathbf{\Gamma}), t = 1, \dots, \mathrm{T}$, $N_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathbf{\Lambda})$, $X_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathrm{A}\mathbf{\Gamma}\mathrm{A}^* + \mathbf{\Lambda})$. Матрицы $\mathbf{\Gamma}$ и $\mathbf{\Lambda}$ предполагаются диагональными, т.е. сигналы не коррелированы между собой, шумы также не коррелированы между собой, корреляция между сигналами и шумами также отсутствует. Составим EM-алгоритм для двух случаев:

- Известный шум;
- Неизвестный шум.

§2 Известный шум

Воспользуемся EM-алгоритмом для того, чтобы определить значения параметров $\Psi = (\theta, \mathbf{\Gamma})$, пропущенные значения $X_m = \{X_{t,m_t}\}_{t=1}^{\mathrm{T}}$ и сигналы S рассматриваются как латентные переменные. Наблюдения X_t, $t = 1, \ldots, \mathrm{T}$ предполагаются независимыми и одинаково распределенными.

Инициализация параметров

Оценим вектор угловых координат источников $\theta^{(0)}$ следующим образом:

1. Выберем число ν, которое будет соответствовать первому компоненту вектора $\theta^{(0)}$:

$$\nu \sim \mathcal{U}([-\pi; \pi]); \tag{3}$$

2. Оценим компоненты вектора $\theta^{(0)}$ так: $\theta_i^{(0)} = (\nu + (i-1) \cdot \frac{2\pi}{\mathrm{K}}) \bmod 2\pi, i = 1, \ldots, \mathrm{K}$. При этом, $a \bmod b = a - b \cdot \lfloor \frac{a}{b} \rfloor$.

Диагональные элементы матрицы $\mathbf{\Gamma}$ задаем с помощью наименьших квадратов:

$$\Gamma = A^{+}(\hat{R}_0 - Q)(A^{+})^{*}. \tag{4}$$

E-шаг

Требуется найти математическое ожидание полного правдоподобия с учетом текущей оценки параметров и апостериорного совместного распределения пропущенных значений в наблюдениях X_m и сигналов S

$$\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[\log P(X,S)]. \tag{5}$$

Преобразуем выражение, обозначив через $\mathcal{I}$ условную информацию $X_o = x_o, \Psi^{(\tau-1)}$:

$$\begin{gathered}
\mathbb{E}_{(X_m,S)|\mathcal{I}}[\log P(X,S)] = \\
\mathbb{E}_{(X_m,S)|\mathcal{I}}[\log(P(X|S)P(S))] = \\
\mathbb{E}_{(X_m,S)|\mathcal{I}}[\log P(X|S) + \log P(S)] = \\
-\mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^{*}]\right) - 2\,\mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^{*}]\right) \\
+\mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^{*}]\mathrm{A}^{*}(\theta)\big) + \log|\mathbf{\Gamma}| + \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^{*}]\big)\Big].
\end{gathered}$$

Для нахождения условных моментов, указанных в формуле выше, требуется найти апостериорное распределение скрытых переменных. Воспользуемся формулой произведения плотностей:

$$P(X_m, S|\mathcal{I}) = P(X_m|\mathcal{I}) \cdot P(S|\mathcal{I}, X_m). \tag{6}$$

Сначала найдем апостериорное распределение $P(X_m|\mathcal{I})$, причем ввиду того, что индексы, соответствующие пропущенным значениям в наблюдениях, могут отличаться в зависимости от t, будем находить апостериорное распределение для каждого X_{t,m_t}. Для достижения этой цели, для каждой пары $\{(o_t, m_t) : m_t \neq \varnothing\}$ создадим разбиение оценки ковариационной матрицы наблюдений $\hat{\mathrm{R}}$ на блоки, индуцированное этим разбиением множества индексов, оно имеет следующий вид:

$$\hat{\mathrm{R}} = \begin{pmatrix} \hat{\mathrm{R}}_{o_t,o_t} & \hat{\mathrm{R}}_{o_t,m_t} \\ \hat{\mathrm{R}}_{m_t,o_t} & \hat{\mathrm{R}}_{m_t,m_t} \end{pmatrix}, \tag{7}$$

где каждый блок определяется как

$$\hat{\mathrm{R}}_{a,b} = (\hat{\mathrm{R}}_{ij})_{i \in a, j \in b}.$$

Для каждого наблюдения, содержащего пропуски, требуется найти апостериорное распределение пропущенных значений, $P(X_{t,m_t}|X_{t,o_t} = x_{t,o_t}, \Psi^{(\tau-1)}), t = 1, \ldots, \mathrm{T}, |m_t| > 0$. Обозначим через $\mathcal{I}_t$ условную информацию $X_{t,o_t} = x_{t,o_t}, \Psi^{(\tau-1)}$. Параметры апостериорного распределения $P(X_{t,m_t}|\mathcal{I}_t), t = 1, \ldots, \mathrm{T}$ на итерации τ можно найти следующим образом:

$$
\begin{cases}
\mu_{X_{t,m_t}|\mathcal{I}_t} = \hat{\mathrm{R}}_{m_t,o_t} \left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1} \cdot x_{t,o_t}, \\
\Sigma_{X_{t,m_t}|\mathcal{I}_t} = \hat{\mathrm{R}}_{m_t,m_t} - \hat{\mathrm{R}}_{m_t,o_t} \left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1} \hat{\mathrm{R}}_{o_t,m_t},
\end{cases}
\tag{8}
$$

где $\hat{\mathrm{R}}_{o_t,o_t} = \hat{\mathrm{R}}^{(\tau-1)}_{o_t,o_t}$, $\hat{\mathrm{R}}_{o_t,m_t} = \hat{\mathrm{R}}^{(\tau-1)}_{o_t,m_t}$, $\hat{\mathrm{R}}_{m_t,o_t} = \hat{\mathrm{R}}^{(\tau-1)}_{m_t,o_t}$, $\hat{\mathrm{R}}_{m_t,m_t} = \hat{\mathrm{R}}^{(\tau-1)}_{m_t,m_t}$.
Для каждого наблюдения X_t оценим условную ковариационную матрицу $\widetilde{\Sigma}_{X_t} = \mathbb{E}_{X_{m_t}|\mathcal{I}_t}[X_t X_t^*]$:

$$
\widetilde{\Sigma}_{X_t} = \begin{pmatrix} \mu_{X_{t,o_t}|\mathcal{I}_t} \cdot \mu^*_{X_{t,o_t}|\mathcal{I}_t} & \mu_{X_{t,o_t}|\mathcal{I}_t} \cdot \mu^*_{X_{t,m_t}|\mathcal{I}_t} \\ \mu_{X_{t,m_t}|\mathcal{I}_t} \cdot \mu^*_{X_{t,o_t}|\mathcal{I}_t} & \mu_{X_{t,m_t}|\mathcal{I}_t} \cdot \mu^*_{X_{t,m_t}|\mathcal{I}_t} + \Sigma_{X_{t,m_t}|\mathcal{I}_t} \end{pmatrix} =
$$
$$
\mathbb{E}[X_t|\mathcal{I}_t] \cdot \mathbb{E}[X_t|\mathcal{I}_t]^* + \begin{pmatrix} \mathbf{O}_{o_t,o_t} & \mathbf{O}_{o_t,m_t} \\ \mathbf{O}_{m_t,o_t} & \Sigma_{X_{t,m_t}|\mathcal{I}_t} \end{pmatrix},
$$

где $\mathbf{O}_{o_t,o_t}$, $\mathbf{O}_{o_t,m_t}$, $\mathbf{O}_{m_t,o_t}$ – нулевые блочные матрицы. Разбиение указанной матрицы на четыре блока, три из которых состоят из нулей, индукцировано разбиением множества индексов $\{1, \ldots, \mathrm{L}\}$ на множества o_t, m_t. Оценим ковариационную матрицу наблюдений с учетом текущей оценки параметров и доступных наблюдений $\widetilde{\Sigma}_X = \mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^*]$ (действуем по аналогии с М-шагом для стандартного EM-алгоритма оценки параметров гауссовского распределения по выборке из одинаково распределенных наблюдений, содержащих пропуски).

$$
\widetilde{\Sigma}_X = \frac{1}{\mathrm{T}} \hat{x}\hat{x}^* + \frac{1}{\mathrm{T}} \Big(\sum_{t=1}^{G} \widetilde{R}_t \Big),
\tag{9}
$$

где $\widetilde{R}_t$ – матрица, в которой элементы $(\widetilde{R}_t)_{ij}$, такие что $i, j \in m_t$ образуют $\Sigma^{(\tau)}_{X_{t,m_t}|X_{t,o_t}=x_{t,o_t},\Psi^{(\tau-1)}}$, а остальные элементы равны нулю, $\hat{x}$ – реализация x матрицы наблюдений X, в которой значения x_{t,m_t} оценены как $\hat{x}_{t,m_t} = \mu^{(\tau)}_{X_{t,m_t}|X_{t,o_t}=x_{t,o_t},\Psi^{(\tau-1)}}$ для всех $t \in \{1, \ldots, \mathrm{T}\}$. Параметры апостериорного распределения $P(S_t|\mathcal{I}_t, X_{t,m_t}), t = 1, \ldots, \mathrm{L}$ можно найти исходя из следующих формул:

$$
\begin{cases}
\mu_{S_t|\mathcal{I}_t,X_{t,m_t}} = \mathbf{\Gamma}\mathrm{A}^* \Big(\widetilde{\Sigma}_X\Big)^{-1} \mathbb{E}[X_t|\mathcal{I}_t], \\
\Sigma_{S|\mathcal{I}_t,X_{t,m_t}} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathrm{A}^* \Big(\widetilde{\Sigma}_X\Big)^{-1} \mathrm{A}\mathbf{\Gamma},
\end{cases}
\tag{10}
$$

где $\mathrm{A} = \mathrm{A}(\theta^{(\tau-1)}), \mathbf{\Gamma} = \mathbf{\Gamma}^{(\tau-1)}$, $\hat{x}_t$ – реализация x_t наблюдения X_t, в котором значения x_{t,m_t} оценены как $\hat{x}_{t,m_t}, t = 1, \ldots, \mathrm{T}$.
Оценим ковариационную матрицу сигналов с учетом текущей оценки параметров и доступных наблюдений $\widetilde{\Sigma}_S = \mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]$.

$$
\widetilde{\Sigma}_S = \frac{1}{\mathrm{T}} \hat{s}\hat{s}^* + \frac{1}{\mathrm{T}} \Big(\sum_{t=1}^{T} \Sigma^{(\tau)}_{S_t|X_{t,o_t}=x_{t,o_t},X_{t,m_t},\Psi^{(\tau-1)}} \Big),
\tag{11}
$$

где $\hat{s} = [\mu^{(\tau)}_{S_1|X_{1,o_1}}, \ldots, \mu^{(\tau)}_{S_\mathrm{L}|X_{\mathrm{L},o_\mathrm{L}}}]$. Заметим, что $\Sigma_{S_t|X_{t,o_t}=x_{t,o_t},X_{t,m_t},\Psi^{(\tau-1)}}$ не зависит от t. Остается оценить кросс-ковариацию $\Sigma_{XS} = \mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*]$:

$$
\begin{aligned}
\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*] &= \mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^*](\hat{\mathrm{R}}^{(\tau-1)})^{-1}\mathrm{A}(\theta^{(\tau-1)})\mathbf{\Gamma}^{(\tau-1)} \\
&= \Sigma_X(\hat{\mathrm{R}}^{(\tau-1)})^{-1}\mathrm{A}(\theta^{(\tau-1)})\mathbf{\Gamma}^{(\tau-1)}.
\end{aligned}
\tag{12}
$$

M-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

$$\Psi^{(\tau)} = \arg\max_{\Psi} \mathbb{E}_{(X_m,S)|\mathcal{I}}[\log P(X,S)] =$$

$$\arg\min_{\Psi} \mathrm{T}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^*]\right) - 2\,\mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*]\right)$$

$$+\,\mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\mathrm{A}^*(\theta)\big) + \log|\mathbf{\Gamma}| + \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\big)\Big].$$

Оценим угловые координаты источников θ:

$$\theta^{(\tau)} = \arg\min_{\theta} \mathcal{J}(\theta) = \arg\min_{\theta} \Big[-2\,\mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\Sigma_{XS}\right)$$

$$+\,\mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\Sigma_S\mathrm{A}^*(\theta)\big)\Big] =$$

$$\arg\min_{\theta} ||\mathbf{\Lambda}^{-1/2}(\Sigma_{XS} - \mathrm{A}(\theta)\Sigma_S)||_F^2.$$

Для решения этой оптимизационной задачи можно использовать сочетание метода последовательного квадратичного программирования и линейного поиска.
Оценим ковариацию сигналов $\mathbf{\Gamma}$:

$$\mathbf{\Gamma}^{(\tau)} = \arg\min_{\mathbf{\Gamma}} \mathcal{K}(\mathbf{\Gamma}|\mathbf{\Gamma}^{(\tau-1)}) = \mathbb{E}_{(X_m,S)|\mathcal{I}}\mathrm{T}\Big[\log|\mathbf{\Gamma}| + \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\big)\Big].$$

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим минимум функции, при этом обозначим через M величину $\mathbb{E}_{(X_m,S)|\mathcal{I}}[S_tS_t^*]$:

$$\frac{\partial}{\partial\mathbf{\Gamma}}\log(\mathrm{Det}(\mathbf{\Gamma})) = \mathbf{\Gamma}^{-1},$$

$$\frac{\partial}{\partial\mathbf{\Gamma}}\mathrm{Tr}\big(\mathbf{\Gamma}^{-1}M\big) = -\mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1},$$

$$\frac{\partial\mathcal{K}(\mathbf{\Gamma})}{\partial\mathbf{\Gamma}} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1}.$$

Приравняем производную к нулю (функция по $\mathbf{\Gamma}$ выпукла):

$$O = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1} \Rightarrow M = \mathrm{T}\mathbf{\Gamma} \Rightarrow \mathbf{\Gamma}^{(\tau)} = \widetilde{\Sigma}_S.$$

Предполагая, что сигналы некоррелированны, будем использовать лишь диагональное приближение матрицы, приравняв элементы вне главной диагонали к нулю:

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D}\Big[\widetilde{\Sigma}_S\Big]. \tag{13}$$

Обновляем оценку ковариации наблюдений с учетом поулченных оценок параметров:

$$\hat{\mathrm{R}}^{(\tau)} = \mathrm{A}(\theta^{(\tau)})\mathbf{\Gamma}\mathrm{A}^*(\theta^{(\tau)}) + \mathbf{\Lambda}. \tag{14}$$

Список источников

1. Dempster, A.P.; Laird, N.M.; Rubin, D.B. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B (Methodol.) 1977, 39, 1–38.