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 — набор наблюдений, полученных сенсорами в моменты времени $t = 1, \ldots, \mathrm{T}$, 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 $o_t \cup m_t = \{1, \ldots, \mathrm{T}\}, o_t \cap m_t = \varnothing, \forall t \in \{1, \ldots, \mathrm{T}\}$. Предполагается, что $\nexists o_t : o_t = \varnothing$, 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

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

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

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

Сигналы, испускаемые источниками, шумы на сенсорах и наблюдения предполагаются стохастическими: $S_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{K}\times 1}, \mathbf{\Gamma})$, $N_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathbf{\Lambda})$, $X_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathrm{A}\mathbf{\Gamma}\mathrm{A}^* + \mathbf{\Lambda})$, $t = 1, \ldots, \mathrm{T}$. 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, $t = 1, \dots, \mathrm{T}$ 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: $\theta_i^{(0)} = (\nu + (i-1) \cdot \frac{2\pi}{\mathrm{K}}) \bmod 2\pi, i = 1, \dots, \mathrm{K}$. При этом, $a \bmod b = a - b \cdot \lfloor \frac{a}{b} \rfloor$.

Начальную оценку ковариационной матрицы наблюдений $\hat{\mathrm{R}}^{(0)}$ получаем на основе доступных данных. Начальную ковариацию сигналов $\Gamma^{(0)}$ получаем на основе метода наименьших квадратов с помощью $\theta^{(0)}, \hat{\mathrm{R}}^{(0)}$.

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

$$\mathbb{E}_{(X_m, S)|X_o = x_o, \Psi^{(\tau-1)}}[\log P(X, S)]. \tag{4}$$

Преобразуем выражение, учитывая тот факт, что наблюдения являются независимыми, и обозначив через $\mathcal{I}$ условную информацию $X_o = x_o, \Psi^{(\tau-1)}$ и через $\mathcal{I}_t$ условную информацию $X_{t,o_t} = x_{t,o_t}, \Psi^{(\tau-1)}$:

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

Выражение $\mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[\cdot]$ соответствуют случаю максимально полного набора латентных переменных: некоторые наблюдения в наборе могут не содержать пропусков, и тогда это выражение упрощается до $\mathbb{E}_{S_t|\mathcal{I}_t}[\cdot]$. Для нахождения условных моментов, указанных в формуле выше, требуется найти апостериорное распределение скрытых переменных.

Случай неполных наблюдений

Воспользуемся формулой произведения плотностей:

$$P(X_{t,m_t}, S_t|\mathcal{I}_t) = P(X_{t,m_t}|\mathcal{I}_t) \cdot P(S_t|X_{t,m_t}, \mathcal{I}_t). \tag{5}$$

Сначала найдем апостериорное распределение $P(X_m|\mathcal{I})$. 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:

$$\hat{\mathrm{R}} = \begin{pmatrix} \hat{\mathrm{R}}_{o_t,o_t} & \hat{\mathrm{R}}_{o_t,m_t} \\ \hat{\mathrm{R}}_{m_t,o_t} & \hat{\mathrm{R}}_{m_t,m_t} \end{pmatrix}, \tag{6}$$

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, $P(X_{t,m_t}|X_{t,o_t} = x_{t,o_t}, \Psi^{(\tau-1)}), t = 1,\ldots,\mathrm{T}, m_t \neq \varnothing$. 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:

$$\begin{cases} \mathbb{E}[X_{t,m_t}|\mathcal{I}_t] = \hat{\mathrm{R}}_{m_t,o_t}\left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1} \cdot x_{t,o_t}, \\ \Sigma_{X_{t,m_t}|\mathcal{I}_t} = \hat{\mathrm{R}}_{m_t,m_t} - \hat{\mathrm{R}}_{m_t,o_t}\left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1}\hat{\mathrm{R}}_{o_t,m_t}, \end{cases} \tag{7}$$

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, содержащего пропуски, оценим условный второй начальный момент $\mathbb{E}[X_tX_t^*|\mathcal{I}_t]$:

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

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.
Параметры апостериорного распределения $P(S_t|\mathcal{I}_t, X_{t,m_t}), t = 1,\ldots,\mathrm{T}$ можно найти исходя из следующих формул, если $m_t \neq \varnothing$:

$$\begin{cases} \mathbb{E}[S_t|\mathcal{I}_t, X_{t,m_t}] = \mathbf{\Gamma}\mathrm{A}^*\hat{\mathrm{R}}^{-1}\mathbb{E}[X_t|\mathcal{I}_t], \\ \Sigma_{S_t|\mathcal{I}_t,X_{t,m_t}} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathrm{A}^*\hat{\mathrm{R}}^{-1}\mathrm{A}\mathbf{\Gamma} + \mathbf{\Gamma}\mathrm{A}^*\hat{\mathrm{R}}^{-1}\mathbb{E}[X_tX_t^*|\mathcal{I}_t]\hat{\mathrm{R}}^{-1}A\mathbf{\Gamma}, \end{cases} \tag{8}$$

где $\mathrm{A} = \mathrm{A}(\theta^{(\tau-1)})$, $\mathbf{\Gamma} = \mathbf{\Gamma}^{(\tau-1)}$, $\hat{\mathrm{R}} = \hat{\mathrm{R}}^{(\tau-1)}$.
Оценим $\mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[S_tS_t^*]$:

$$\mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[S_tS_t^*] = \mathbb{E}[S_t|\mathcal{I}_t, X_{t,m_t}]\cdot\mathbb{E}[S_t|\mathcal{I}_t, X_{t,m_t}]^* + \Sigma_{S_t|\mathcal{I}_t,X_{t,m_t}}, \tag{9}$$

Оценим $\mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[X_tS_t^*]$:

$$\mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[X_tS_t^*] = \mathbb{E}[X_tX_t^*|\mathcal{I}_t]\hat{\mathrm{R}}^{-1}\mathrm{A}\mathbf{\Gamma} + \mathbb{E}[X_t|\mathcal{I}_t]\cdot\mathbb{E}[S_t|\mathcal{I}_t, X_{t,m_t}]^*, \tag{10}$$

где $\mathrm{A} = \mathrm{A}(\theta^{(\tau-1)})$, $\hat{\mathrm{R}} = \hat{\mathrm{R}}^{(\tau-1)}$, $\mathbf{\Gamma} = \mathbf{\Gamma}^{(\tau-1)}$.

Случай полных наблюдений

Если $m_t = \varnothing$, то апостериорное распределение упрощается следующим образом: $P(X_{t,m_t}, S_t|\mathcal{I}_t) = P(S_t|\mathcal{I}_t)$. Параметры апостериорного распределения $P(S_t|\mathcal{I}_t), t = 1,\ldots,\mathrm{T}$:

$$\begin{cases} \mathbb{E}[S_t|\mathcal{I}_t] = \mathbf{\Gamma}\mathrm{A}^*\hat{\mathrm{R}}^{-1}\mathbb{E}[X_t|\mathcal{I}_t], \\ \Sigma_{S_t|\mathcal{I}_t} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathrm{A}^*\hat{\mathrm{R}}^{-1}\mathrm{A}\mathbf{\Gamma}. \end{cases} \tag{11}$$

При этом:

$$\mathbb{E}_{S_t|\mathcal{I}_t}[X_t X_t^*] = \mathbb{E}[X_t|\mathcal{I}_t] \cdot \mathbb{E}[X_t|\mathcal{I}_t]^* = x_t x_t^*, \tag{12}$$

$$\mathbb{E}_{S_t|\mathcal{I}_t}[X_t S_t^*] = \mathbb{E}[X_t|\mathcal{I}_t] \cdot \mathbb{E}[S_t|\mathcal{I}_t]^* = x_t \cdot \mathbb{E}[S_t|\mathcal{I}_t]^*, \tag{13}$$

$$\mathbb{E}_{S_t|\mathcal{I}_t}[S_t S_t^*] = \mathbb{E}[S_t|\mathcal{I}_t] \cdot \mathbb{E}[S_t|\mathcal{I}_t]^* + \Sigma_{S_t|\mathcal{I}_t}. \tag{14}$$

Агрегация

Теперь оценим вторые начальные моменты $\mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^*]$, $\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]$, $\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*]$:

$$\mathbb{E}_{(X_m,S)|\mathcal{I}}[XX^*] = \frac{1}{\mathrm{T}} \sum_{t=1}^{T} \mathbb{E}[X_t X_t^* |\mathcal{I}_t], \tag{15}$$

$$\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*] = \frac{1}{\mathrm{T}} \Big[\sum_{t\in T_1} \mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[S_t S_t^*] + \sum_{t\in T_2} \mathbb{E}_{S_t|\mathcal{I}_t}[S_t S_t^*] \Big], \tag{16}$$

$$\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*] = \frac{1}{\mathrm{T}} \Big[\sum_{t\in T_1} \mathbb{E}_{(X_{t,m_t},S_t)|\mathcal{I}_t}[X_t S_t^*] + \sum_{t\in T_2} \mathbb{E}_{S_t|\mathcal{I}_t}[X_t S_t^*] \Big], \tag{17}$$

где $T_1 = \{y \in \mathbb{N} : m_y \neq \varnothing, 1 \leq y \leq T\}$, $T_2 = \{y \in \mathbb{N} : m_y = \varnothing, 1 \leq y \leq T\}$.

М-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:

$$\theta^{(\tau)} = \arg\min_{\theta} \mathcal{J}(\theta|\theta^{(\tau-1)}) = \arg\min_{\theta} \Big[-2\,\mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*]\right)$$
$$+ \mathrm{Tr}\big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\mathrm{A}^*(\theta)\big)\Big] =$$
$$\arg\min_{\theta} ||\mathbf{\Lambda}^{-1/2}(\mathbb{E}_{(X_m,S)|\mathcal{I}}[XS^*] - \mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*])||_F^2.$$

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

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

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим минимум функции:

$$\frac{\partial}{\partial\mathbf{\Gamma}} \log(\mathrm{Det}(\mathbf{\Gamma})) = \mathbf{\Gamma}^{-1},$$
$$\frac{\partial}{\partial\mathbf{\Gamma}} \mathrm{Tr}\big(\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\big) = -\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\mathbf{\Gamma}^{-1},$$
$$\frac{\partial\mathcal{K}(\mathbf{\Gamma})}{\partial\mathbf{\Gamma}} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\mathbf{\Gamma}^{-1}.$$

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

$$O = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\mathbf{\Gamma}^{-1} \Rightarrow \mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*] = \mathbf{\Gamma} \Rightarrow \mathbf{\Gamma}^{(\tau)} = \mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*].$$

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

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D}\Big[\mathbb{E}_{(X_m,S)|\mathcal{I}}[SS^*]\Big]. \tag{18}$$

Обновляем оценку ковариации наблюдений с учетом полученных оценок параметров:

$$\hat{\mathrm{R}}^{(\tau)} = \mathrm{A}(\theta^{(\tau)})\mathbf{\Gamma}\mathrm{A}^*(\theta^{(\tau)}) + \mathbf{\Lambda}. \tag{19}$$

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.