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 G снимков полученного сигнала, причем ввиду технических неполадок, 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 $t = 1, \dots, \mathrm{G}$, 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 $X_o = \{X_{t,o_t}\}_{t=1}^{\mathrm{G}}$ и ненаблюдаемой: $X_m = \{X_{t,m_t}\}_{t=1}^{\mathrm{G}}$, причем $o_t \cup m_t = \{1, \dots, \mathrm{L}\}, o_t \cap m_t = \varnothing, \forall t \in \{1, \dots, \mathrm{G}\}$. 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 $N = \{N_t\}_{t=1}^{\mathrm{G}}$ соответствует набору шумов, связанных с датчиками в моменты времени $t = 1, \dots, \mathrm{G}$, $S = \{S_t\}_{t=1}^{\mathrm{G}}$ – соответствует набору сигналов, испускаемых источниками в моменты времени $t = 1, \dots, \mathrm{G}$, 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: $S_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{K}\times 1}, \mathbf{\Gamma}), t = 1, \dots, \mathrm{G}$, $N_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathbf{\Lambda})$, $X_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathrm{A}\mathbf{\Gamma}\mathrm{A}^* + \mathbf{\Lambda})$. Матрицы $\mathbf{\Gamma}$ и $\mathbf{\Lambda}$ предполагаются диагональными, т.е. и сигналы не коррелированы между собой, шумы также не коррелированы между собой. 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 $X_m = \{X_{t,m_t}\}_{t=1}^{\mathrm{G}}$ рассматриваются как латентные переменные. Наблюдения X_t, $t = 1, \ldots, \mathrm{G}$ 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 $\theta^{(0)}$:
$$\nu \sim \mathcal{U}([-\pi; \pi]); \tag{2}$$

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 равномерного распределения:
$$p_{jj} \sim \mathcal{U}([0,2; 5]), \tag{3}$$
где $j = 1, \ldots, \mathrm{K}$.

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

Для нахождения условных моментов, указанных в формуле выше, требуется найти апостериорное распределение скрытых переменных. Воспользуемся формулой произведения плотностей:
$$P(X_m, S|X_o = x_o, \Psi^{(\tau-1)}) = P(X_m|X_o = x_o, \Psi^{(\tau-1)}) \cdot P(S|X_o = x_o, X_m, \Psi^{(\tau-1)}). \tag{5}$$

Сначала найдем апостериорное распределение $P(X_m|X_o = x_o, \Psi^{(\tau-1)})$, 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:
$$\hat{\mathrm{R}} = \begin{pmatrix} \hat{\mathrm{R}}_{o_t,o_t} & \hat{\mathrm{R}}_{o_t,m_t} \\ \hat{\mathrm{R}}_{m_t,o_t} & \hat{\mathrm{R}}_{m_t,m_t} \end{pmatrix}, \tag{6}$$
где каждый блок определяется как
$$\hat{\mathrm{R}}_{a,b} = (\hat{\mathrm{R}}_{ij})_{i\in a, j\in b}.$$

Параметры апостериорного распределения $P(X_{t,m_t}|X_{t,o_t} = x_{t,o_t}, \Psi^{(\tau-1)}), t = 1, \ldots, \mathrm{G}$ на итерации τ можно найти следующим образом:

$$\begin{cases} \mu^{(\tau)}_{X_{t,m_t}|X_{t,o_t}=x_{t,o_t},\Psi^{(\tau-1)}} = \hat{\mathrm{R}}_{m_t,o_t}\left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1} \cdot x_{t,o_t}, \\ \Sigma^{(\tau)}_{X_{t,m_t}|X_{t,o_t}=x_{t,o_t},\Psi^{(\tau-1)}} = \hat{\mathrm{R}}_{m_t,m_t} - \hat{\mathrm{R}}_{m_t,o_t}\left(\hat{\mathrm{R}}_{o_t,o_t}\right)^{-1}\hat{\mathrm{R}}_{o_t,m_t}, \end{cases} \tag{7}$$

где $\hat{\mathrm{R}}_{o_t,o_t} = \hat{\mathrm{R}}^{(\tau-1)}_{o_t,o_t}$, $\hat{\mathrm{R}}_{o_t,m_t} = \hat{\mathrm{R}}^{(\tau-1)}_{o_t,m_t}$, $\hat{\mathrm{R}}_{m_t,o_t} = \hat{\mathrm{R}}^{(\tau-1)}_{m_t,o_t}$, $\hat{\mathrm{R}}_{m_t,m_t} = \hat{\mathrm{R}}^{(\tau-1)}_{m_t,m_t}$.
Оценим ковариационную матрицу наблюдений с учетом текущей оценки параметров и доступных наблюдений $\widetilde{\Sigma}_X = \mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XX^*]$ (действуем по аналогии с М-шагом для стандартного ЕМ-алгоритма оценки параметров гауссовского распределения по выборке из одинаково распределенных наблюдений, содержащих пропуски).

$$\widetilde{\Sigma}_X = \frac{1}{\mathrm{G}}\hat{x}\hat{x}^* + \frac{1}{\mathrm{G}}\Big(\sum_{t=1}^{G}\widetilde{R}_t\Big), \tag{8}$$

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 $t \in \{1, \ldots, \mathrm{G}\}$. Параметры апостериорного распределения $P(S_t|X_{t,o_t} = x_{t,o_t}, X_{t,m_t}, \Psi^{(\tau-1)}), t = 1, \ldots, \mathrm{L}$ можно найти исходя из следующих формул:

$$\begin{cases} \mu^{(\tau)}_{S_t|X^{(\tau)}_{t,o_t}=x_{t,o_t},X_{t,m_t},\Psi^{(\tau-1)}} = \mathbf{\Gamma}\mathrm{A}^*\left(\widetilde{\Sigma}_X\right)^{-1}\hat{x}_t, \\ \Sigma^{(\tau)}_{S|X^{(\tau)}_{t,o_t}=x_{t,o_t},X_{t,m_t},\Psi^{(\tau-1)}} = \mathbf{\Gamma} - \mathbf{\Gamma}\mathrm{A}^*\left(\widetilde{\Sigma}_X\right)^{-1}\mathrm{A}\mathbf{\Gamma}, \end{cases} \tag{9}$$

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 $\hat{x}_{t,m_t}, t = 1, \ldots, \mathrm{G}$.
Оценим ковариационную матрицу сигналов с учетом текущей оценки параметров и доступных наблюдений $\widetilde{\Sigma}_S = \mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[SS^*]$.

$$\widetilde{\Sigma}_S = \frac{1}{\mathrm{G}}\hat{s}\hat{s}^* + \frac{1}{\mathrm{G}}\Big(\sum_{t=1}^{G}\Sigma^{(\tau)}_{S_t|X_{t,o_t}=x_{t,o_t},X_{t,m_t},\Psi^{(\tau-1)}}\Big), \tag{10}$$

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 $\Sigma_{XS} = \mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XS^*]$:

$$\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XS^*] = \mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XX^*](\hat{\mathrm{R}}^{(\tau-1)})^{-1}\mathrm{A}(\theta^{(\tau-1)})\mathbf{\Gamma}^{(\tau-1)} = \Sigma_X(\hat{\mathrm{R}}^{(\tau-1)})^{-1}\mathrm{A}(\theta^{(\tau-1)})\mathbf{\Gamma}^{(\tau} \tag{11}$$

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

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

$$\arg\min_{\Psi}\mathrm{G}\Big[\log|\mathbf{\Lambda}| + \mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XX^*]\right) - 2\,\mathrm{Tr}\left(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[XS^*]\right)$$
$$+\,\mathrm{Tr}\Big(\mathbf{\Lambda}^{-1}\mathrm{A}(\theta)\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[SS^*]\mathrm{A}(\theta)^*\Big) + \log|\mathbf{\Gamma}| + \mathrm{Tr}\Big(\mathbf{\Gamma}^{-1}\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[SS^*]\Big)\Big].$$

Оценим угловые координаты источников θ:

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

Оценим ковариацию сигналов $\mathbf{\Gamma}$:

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

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим минимум функции, при этом обозначим через M величину $\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[S_tS_t^*]$:

$$\frac{\partial}{\partial\mathbf{\Gamma}}\log(\mathrm{Det}(\mathbf{\Gamma})) = \mathbf{\Gamma}^{-1},$$
$$\frac{\partial}{\partial\mathbf{\Gamma}}\mathrm{Tr}\big(\mathbf{\Gamma}^{-1}M\big) = -\mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1},$$
$$\frac{\partial\mathcal{K}(\mathbf{\Gamma})}{\partial\mathbf{\Gamma}} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1}.$$

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

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

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

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D}\Big[\widetilde{\Sigma}_S\Big]. \tag{12}$$

Обновляем оценку ковариации наблюдений:

$$\hat{\mathrm{R}}^{(\tau)} = \mathrm{A}(\theta^{(\tau)})\mathbf{\Gamma}[\mathrm{A}(\theta^{(\tau)})]^* + \mathbf{\Lambda}. \tag{13}$$

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.