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 $K < L$. Этим источникам соответствуют угловые координаты (DoA) θ, практически не изменяющийся 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 $t = 1, 2, ..., \mathrm{G}$, 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 $o_t \cup m_t = \{1, ..., L\}, o_t \cap m_t = \varnothing, \forall t \in \{1, ..., \mathrm{G}\}$. 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 $t = 1, 2, ..., \mathrm{G}$, $S = \{S_t\}_{t=1}^{\mathrm{G}}$ – соответствует набору сигналов, испускаемых источниками в моменты времени $t = 1, 2, ..., \mathrm{G}$, 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, также как и шумы на сенсорах, предполагаются стохастическими: $S_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{K}\times 1}, \mathbf{\Gamma}), t = 1, 2, ..., \mathrm{G}$, $N_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathbf{\Lambda})$. Матрицы $\mathbf{\Gamma}$ и $\mathbf{\Lambda}$ предполагаются диагональными, т.е. и сигналы, и шумы, являются некоррелированными. Для простоты дальнейших рассуждений введем также следующие величины:

- L_{o_t} — число исправных сенсоров в момент времени t;
- L_{m_t} — число неисправных сенсоров в момент времени t.

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, $t = 1, 2, ..., \mathrm{G}$ 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: $\theta_i^{(0)} = (\nu + (i-1) \cdot \frac{2\pi}{\mathrm{K}}) \bmod 2\pi, i = 1, 2, ..., \mathrm{K}$. 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 $j = 1, 2, ..., \mathrm{K}$.

E-шаг

Требуется найти условное математическое ожидание с учетом текущей оценки параметров и апостериорного совместного распределения ненаблюдаемых/пропущенных принятых сигналов X_m и исходных сигналов S
$$\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[\log P(X, S)]. \tag{4}$$
Сначала найдем апостериорное распределение $P(X_m|X_o = x_o, \Psi)$, воспользуемся формулой произведения плотностей:
$$P((X_m, S)|X_o = x_o, \Psi) = P(X_m|X_o = x_o, \Psi) \cdot P(S|X_o = x_o, X_m = \hat{x}_m, \Psi), \tag{5}$$
где $\hat{x}_m$ – оценка ненаблюдаемой части реализации набора наблюдений x, полученная на основе апостериорного распределения $P(X_m|X_o = x_o, \Psi)$.
$$\begin{gathered} X_t = \mathrm{A}S_t + N_t, \\ S_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{K}\times 1}, \mathbf{\Gamma}), \\ X_t \sim \mathcal{CN}(\mathbf{O}_{\mathrm{L}\times 1}, \mathrm{A}\mathbf{\Gamma}\mathrm{A}^* + \mathbf{\Lambda}), \\ X_t|S_t \sim \mathcal{CN}(AS_t, \mathbf{\Lambda}), \end{gathered}$$

$$P(S|\Psi) = \prod_{t=1}^{\mathrm{G}} \frac{1}{\pi^{\mathrm{K}} \operatorname{Det}(\mathbf{\Lambda})} e^{-S_t^*(\mathbf{\Gamma})^{-1}S_t}, \tag{6}$$

$$P(X|\Psi) = \prod_{t=1}^{\mathrm{G}} \frac{1}{\pi^{L} \operatorname{Det}(\mathrm{A}\mathbf{\Gamma}\mathrm{A}^* + \mathbf{\Lambda})} e^{-X_t^*(\mathrm{A}\mathbf{\Gamma}\mathrm{A}^*+\mathbf{\Lambda})^{-1}X_t}, \tag{7}$$

$$P(X|S, \Psi) = \prod_{t=1}^{\mathrm{G}} \frac{1}{\pi^{L} \operatorname{Det}(\mathbf{\Lambda})} e^{-(X_t-\mathrm{A}S_t)^*(\mathbf{\Lambda})^{-1}(X_t-\mathrm{A}S_t)}. \tag{8}$$

Параметры апостериорного распределения $P(X_{t,m_t}|X_{t,o_t}=X_{t,o_t},\Psi)$ на итерации τ можно найти следующим образом:

$$\begin{cases} \mu^{(\tau)}_{X_{t,m_t}|X_{o,t}=X_{t,o_t}} = \hat{R}^{(\tau)}_{X_{t,m_t},X_{t,o_t}}(\hat{R}^{(\tau)}_{X_{t,o_t},X_{t,o_t}})^{-1}\cdot X_{t,o_t}, \\ \Sigma^{(\tau)}_{X_{t,m_t}|X_{o,t}=X_{t,o_t}} = \hat{R}^{(\tau)}_{X_{t,m_t},X_{t,m_t}} - \hat{R}^{(\tau)}_{X_{t,m_t},X_{t,o_t}}(\hat{R}^{(\tau)}_{X_{t,o_t},X_{t,o_t}})^{-1}\hat{R}^{(\tau)}_{X_{t,o_t},X_{t,m_t}}. \end{cases} \tag{9}$$

Пропущенные (ненаблюдаемые) значения X_{t,m_t} оценим через $\hat{x}_{t,m_t}=\mu^{(\tau)}_{X_{t,m_t}|X_{o,t}=X_{t,o_t}}$. Пусть $\hat{x}^{(\tau)}_t$ — вектор x_t, в котором пропущенные значения X_{t,m_t} оценены с помощью $\hat{x}^{(\tau)}_{t,m_t}$, $\hat{x}^{(\tau)}$ — реализация матрицы наблюдений x, в которой пропущенные значения X_{t,m_t} оценены с помощью $\hat{x}^{(\tau)}_{m_t}$ для всех $t\in\{1,2,...,\mathrm{G}\}$.
Параметры апостериорного распределения $P(S_t|X_{t,o_t}=X_{t,o_t},X_{t,m_t}=\hat{x}_{t,m_t},\Psi)$ можно найти исходя из следующих формул

$$\begin{cases} \mu^{(\tau)}_{S_t|X_{t,o_t}=X_{t,o_t},X_{t,m_t}=\hat{x}^{(\tau)}_{t,m_t},\Psi} = \mathbf{\Gamma}\mathrm{A}^*\left(\mathrm{A}\mathbf{\Gamma}\mathrm{A}^*+\mathbf{\Lambda}\right)^{-1}\hat{x}^{(\tau)}_t, \\ \Sigma^{(\tau)}_{S_t|X_{t,o_t}=X_{t,o_t},X_{t,m_t}=\hat{x}^{(\tau)}_{t,m_t},\Psi} = \mathbf{\Gamma}-\mathbf{\Gamma}\mathrm{A}^*\left(\mathrm{A}\mathbf{\Gamma}\mathrm{A}^*+\mathbf{\Lambda}\right)^{-1}\mathrm{A}\mathbf{\Gamma}, \end{cases} \tag{10}$$

где $\mathrm{A}=\mathrm{A}^{(\tau-1)}$, $\mathbf{\Gamma}=\mathbf{\Gamma}^{(\tau-1)}$.
Оценим реализации сигналов через математическое ожидание апостериорного распределения: $\hat{s}^{(\tau)}_t=\mu_{S_t|X_{t,o_t}=X_{t,o_t},X_{t,m_t}}$, $t\in\{1,2,...,\mathrm{G}\}$. Тогда $\hat{s}^{(\tau)}=\left\{\hat{s}^{(\tau)}_t\right\}_{t=1}^{\mathrm{G}}$. Заметим, что $\Sigma_{S_t|X_{t,o_t}=X_{t,o_t},X_{t,m_t}=\hat{x}^{(\tau)}_{t,m_t}}$ не зависит от t. Вернемся к ранее рассмотренному условному математическому ожиданию:

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

Его следует максимизировать, мы можем перейти от логарифма произведения к сумме логарифмов:

$$\log P(X,S|\theta,\mathbf{\Gamma})=\log P(X|S=s,\theta)+\log P(S|\mathbf{\Gamma})=$$
$$=-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Lambda}))-\sum_{t=1}^{\mathrm{G}}(X_t-\mathrm{A}(\theta)S_t)^*\mathbf{\Lambda}^{-1}(X_t-\mathrm{A}(\theta)S_t)-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Gamma}))-\sum_{t=1}^{\mathrm{G}}S_t^*\mathbf{\Gamma}^{-1}S_t.$$

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

$$\Psi^{(\tau)}=\arg\max_{\Psi}\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}[\log P(X,S)]=$$
$$\arg\max_{\Psi}\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}\Bigg[-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Lambda}))-\sum_{t=1}^{\mathrm{G}}(X_t-\mathrm{A}(\theta)S_t)^*\mathbf{\Lambda}^{-1}(X_t-\mathrm{A}(\theta)S_t)$$
$$-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Gamma}))-\sum_{t=1}^{\mathrm{G}}S_t^*\mathbf{\Gamma}^{-1}S_t\Bigg].$$

Оценим угловые координаты источников θ:

$$\theta^{(\tau)}=\arg\max_{\theta}Q(\theta|\theta^{(\tau-1)})=$$
$$\arg\max_{\theta}\mathbb{E}_{(X_m,S)|X_o=x_o,\Psi^{(\tau-1)}}\Bigg[-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Lambda}))-\sum_{t=1}^{\mathrm{G}}(X_t-\mathrm{A}(\theta)S_t)^*\mathbf{\Lambda}^{-1}(X_t-\mathrm{A}(\theta)S_t)$$
$$-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Gamma}))-\sum_{t=1}^{\mathrm{G}}S_t^*\mathbf{\Gamma}^{-1}S_t\Bigg].$$

Тогда максимизируемая функция примет следующий вид:

$$\mathcal{J}(\theta) = \sum_{t=1}^{G} \mathbb{E}\Big[- (X_t - \mathrm{A}(\theta)S_t)^* \mathbf{\Lambda}^{-1}(X_t - \mathrm{A}(\theta)S_t)|X_o = x_o\Big] =$$
$$= -||\mathbf{\Lambda}^{-1/2}(\hat{x} - \mathrm{A}(\theta)\hat{s})||_F^2,$$

где $\hat{x} = \hat{x}^{(\tau)}, \hat{s} = \hat{s}^{(\tau)}$. Оценка УМО, полученная выше, была выведена в документе для детерминированной модели сигналов.

$$\theta^{(\tau)} = \arg\min_{\theta} ||\mathbf{\Lambda}^{-1/2}(\hat{x}^{(\tau)} - \mathrm{A}\hat{s})||_F^2. \tag{11}$$

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

$$\mathbf{\Gamma}^{(\tau)} = \arg\max_{\mathbf{\Gamma}} Q(\mathbf{\Gamma}|\mathbf{\Gamma}^{(\tau-1)}).$$

Пользуемся тем фактом, что полное правдоподобие раскладывается на сумму $\log P(X|S = s) + \log P(S)$. Первый логарифм не зависит от $\mathbf{\Gamma}$. Поэтому максимизируем условное математическое ожидание для $\log P(S|\Psi)$:

$$\mathcal{K}(\mathbf{\Gamma}) = \mathbb{E}_{S|X=\hat{x},\Psi^{(\tau-1)}}\left[-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Gamma})) - \sum_{t=1}^{\mathrm{G}} S_t^*\mathbf{\Gamma}^{-1}S_t\right] =$$
$$-\mathrm{G}\log(\mathrm{Det}(\pi\mathbf{\Gamma})) - \mathbb{E}_{\Psi^{(\tau-1)}}\left[\sum_{t=1}^{\mathrm{G}} S_t^*\mathbf{\Gamma}^{-1}S_t\Big|X = \hat{x}^{(\tau)}\right] =$$
$$-\mathrm{G}\log(\mathrm{Det}(\pi)) - \mathrm{G}\log(\mathrm{Det}(\mathbf{\Gamma})) - \mathbb{E}_{\Psi^{(\tau-1)}}\left[\sum_{t=1}^{\mathrm{G}} S_t^*\mathbf{\Gamma}^{-1}S_t\Big|X = \hat{x}^{(\tau)}\right] =$$
$$-\mathrm{G}\log(\mathrm{Det}(\mathbf{\Gamma})) - \sum_{t=1}^{\mathrm{G}} \mathrm{Tr}\left(\mathbf{\Gamma}^{-1}\mathbb{E}[S_tS_t^*|X = \hat{x}^{(\tau)}]\right)$$

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим максимум функции, при этом обозначим через M величину $\mathbb{E}[S_tS_t^*|X = \hat{x}^{(\tau)}]$:

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

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

$$O = -\mathrm{G}\mathbf{\Gamma}^{-1} + \mathbf{\Gamma}^{-1}M\mathbf{\Gamma}^{-1} \Rightarrow M = \mathrm{G}\mathbf{\Gamma} \Rightarrow \mathbf{\Gamma}^{(\tau)} = \frac{1}{\mathrm{G}}\sum_{t=1}^{\mathrm{G}}\Big(\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^*\Big),$$

и находим аналитическое решение данного уравнения:

$$\mathbf{\Gamma}^{(\tau)} = \frac{1}{\mathrm{G}}\sum_{t=1}^{\mathrm{G}}\Big[\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^*\Big]. \tag{12}$$

Учтем, что $\sum_{t=1}^{\mathrm{G}} a_ta_t^* = \mathrm{AA}^*$, если A – матрица, составленная из столбцов $a_t, t = 1, 2, ..., \mathrm{G}$. Также учтем, что $\Sigma_{S_t|X_t}$ принимает одинаковое значение при любом t:

$$\mathbf{\Gamma}^{(\tau)} = \frac{1}{\mathrm{G}}\hat{s}^{(\tau)} \cdot [\hat{s}^{(\tau)}]^* + \Sigma_{S_1|X_1}. \tag{13}$$

Предполагая, что сигналы некоррелированны, изменим оценку так, чтобы учесть лишь диагональные элементы:

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D}\left[\frac{1}{\mathrm{G}} M_{S|X} \cdot M_{S|X}^{*} + \Sigma_{S_t|X_t}\right]. \tag{14}$$

Обновляем оценку ковариации наблюдений:

$$\hat{R}^{(\tau)} = \mathrm{A}(\theta^{(\tau)})\mathbf{\Gamma}[\mathrm{A}(\theta^{(\tau)})]^{*} + \mathbf{\Lambda}. \tag{15}$$

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.