

ЕМ-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 $t = 1, 2, \dots, G$, 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 $X_o = \{X_{t,o_t}\}_{t=1}^G$ и ненаблюдаемой: $X_m = \{X_{t,m_t}\}_{t=1}^G$, причем $o_t \cup m_t = \{1, \dots, L\}$, $o_t \cap m_t = \emptyset, \forall t \in \{1, \dots, G\}$. Предполагается, что $\#o_t : |o_t| = 0$, т.е. нет таких наблюдений, которые состоят лишь из ненаблюдаемой части. Набор наблюдений X является результатом следующей модели наблюдений:

$$X = AS + N, \quad (1)$$

где $N = \{N_t\}_{t=1}^G$ соответствует набору шумов, связанных с датчиками в моменты времени $t = 1, 2, \dots, G$, $S = \{S_t\}_{t=1}^G$ — соответствует набору сигналов, испускаемых источниками в моменты времени $t = 1, 2, \dots, G$, A — матрица векторов направленности для равномерного линейного массива:

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

Сигналы, испускаемые источниками, также как и шумы на сенсорах, предполагаются стохастическими: $S_t \sim \mathcal{CN}(\mathbf{O}_{K \times 1}, \mathbf{\Gamma})$, $t = 1, 2, \dots, G$, $N_t \sim \mathcal{CN}(\mathbf{O}_{L \times 1}, \mathbf{\Lambda})$. 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 ЕМ-алгоритм для двух случаев:

- Известный шум;
- Неизвестный шум.

§2 Известный шум

Воспользуемся ЕМ-алгоритмом для того, чтобы определить значения параметров $\Psi = (\theta, \Gamma)$, пропущенные значения $X_m = \{X_{t,m_t}\}_{t=1}^G$ рассматриваются как латентные переменные. Наблюдения $X_t, t = 1, 2, \dots, G$ 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 $\theta^{(0)}$:

$$\nu \sim \mathcal{U}([- \pi; \pi]); \quad (2)$$

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

Диагональные элементы матрицы Γ задаем с помощью равномерного распределения:

$$p_{jj} \sim \mathcal{U}([0, 2; 5]), \quad (3)$$

где $j = 1, 2, \dots, K$.

Е-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

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

Сначала найдем апостериорное распределение $P(X_m | X_o = x_o, \Psi)$, воспользуемся формулой произведения плотностей:

$$P((X_m, S) | X_o = x_o, \Psi) = P(X_m | X_o = x_o, \Psi) \cdot P(S | X_o = x_o, X_m = \hat{x}_m, \Psi), \quad (5)$$

где \hat{x}_m – оценка ненаблюдаемой части реализации набора наблюдений x , полученная на основе апостериорного распределения $P(X_m | X_o = x_o, \Psi)$.

$$\begin{aligned} X_t &= AS_t + N_t, \\ S_t &\sim \mathcal{CN}(\mathbf{O}_{K \times 1}, \Gamma), \\ X_t &\sim \mathcal{CN}(\mathbf{O}_{L \times 1}, A\Gamma A^* + \Lambda), \\ X_t | S_t &\sim \mathcal{CN}(AS_t, \Lambda), \end{aligned}$$

$$P(S | \Psi) = \prod_{t=1}^G \frac{1}{\pi^K \text{Det}(\Gamma)} e^{-S_t^* (\Gamma)^{-1} S_t}, \quad (6)$$

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

$$P(X | S, \Psi) = \prod_{t=1}^G \frac{1}{\pi^L \text{Det}(\Lambda)} e^{-(X_t - AS_t)^* (\Lambda)^{-1} (X_t - AS_t)}. \quad (8)$$

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

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

Пропущенные (ненаблюдаемые) значения X_{t,m_t} оценим через $\hat{x}_{t,m_t} = \mu_{X_{t,m_t}|X_{o,t}=X_{t,o_t}}^{(\tau)}$. Пусть $\hat{x}_t^{(\tau)}$ — вектор x_t , в котором пропущенные значения X_{t,m_t} оценены с помощью $\hat{x}_{t,m_t}^{(\tau)}$, $\hat{x}^{(\tau)}$ — реализация матрицы наблюдений x , в которой пропущенные значения X_{t,m_t} оценены с помощью $\hat{x}_{m_t}^{(\tau)}$ для всех $t \in \{1, 2, \dots, G\}$.

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:

$$\begin{cases} \mu_{S_t|X_{t,o_t}=X_{t,o_t}, X_{t,m_t}=\hat{x}_{t,m_t}, \Psi}^{(\tau)} = \mathbf{A}^* (\mathbf{A} \mathbf{A}^* + \mathbf{\Lambda})^{-1} \hat{x}_t^{(\tau)}, \\ \Sigma_{S_t|X_{t,o_t}=X_{t,o_t}, X_{t,m_t}=\hat{x}_{t,m_t}, \Psi}^{(\tau)} = \mathbf{\Gamma} - \mathbf{A}^* (\mathbf{A} \mathbf{A}^* + \mathbf{\Lambda})^{-1} \mathbf{A} \mathbf{\Gamma}, \end{cases} \quad (10)$$

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

Оценим реализации сигналов через математическое ожидание апостериорного распределения: $\hat{s}_t^{(\tau)} = \mu_{S_t|X_{t,o_t}=X_{t,o_t}, X_{t,m_t}=\hat{x}_{t,m_t}, \Psi}^{(\tau)}$, $t \in \{1, 2, \dots, G\}$. Тогда $\hat{s}^{(\tau)} = \{\hat{s}_t^{(\tau)}\}_{t=1}^G$. Заметим, что $\Sigma_{S_t|X_{t,o_t}=X_{t,o_t}, X_{t,m_t}=\hat{x}_{t,m_t}, \Psi}^{(\tau)}$ 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:

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

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

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

Оценим угловые координаты источников θ :

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

Тогда максимизируемая функция примет следующий вид:

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

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

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

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 $\log P(S|\Psi)$:

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

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим максимум функции, при этом обозначим через M величину $\sum_{t=1}^G \mathbb{E}[S_t S_t^* | X = \hat{x}^{(\tau)}]$:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{\Gamma}} \log(\text{Det}(\mathbf{\Gamma})) &= \mathbf{\Gamma}^{-1}, \\ \frac{\partial}{\partial \mathbf{\Gamma}} \text{Tr}(\mathbf{\Gamma}^{-1} M) &= -\mathbf{\Gamma}^{-1} M \mathbf{\Gamma}^{-1}, \\ \frac{\partial \mathcal{K}(\mathbf{\Gamma})}{\partial \mathbf{\Gamma}} &= -G \mathbf{\Gamma}^{-1} + \mathbf{\Gamma}^{-1} M \mathbf{\Gamma}^{-1}.\end{aligned}$$

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

$$0 = -G \mathbf{\Gamma}^{-1} + \mathbf{\Gamma}^{-1} M \mathbf{\Gamma}^{-1} \Rightarrow M = G \mathbf{\Gamma} \Rightarrow \mathbf{\Gamma}^{(\tau)} = \frac{1}{G} \sum_{t=1}^G \left(\Sigma_{S_t|X_t} + \mu_{S_t|X_t} (\mu_{S_t|X_t})^* \right),$$

и находим аналитическое решение данного уравнения:

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

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

$$\mathbf{\Gamma}^{(\tau)} = \frac{1}{G} \hat{s}^{(\tau)} \cdot [\hat{s}^{(\tau)}]^* + \Sigma_{S_t|X_t}. \quad (13)$$

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

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D} \left[\frac{1}{G} \hat{s}^{(\tau)} \cdot [\hat{s}^{(\tau)}]^* + \Sigma_{S_t|X_t} \right]. \quad (14)$$

Обновляем оценку ковариации наблюдений:

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

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.