

ЕСМ-алгоритм, стохастическая модель сигнала

§1 Постановка проблемы

Предположим, у нас имеется линейная антенная решетка, состоящая из L сенсоров. Решетка принимает волны, направленные из M различных источников. Этим источникам соответствует вектор углов прибытия (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 $t = \overline{1, G}$), 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 $X_o = \{X_{o_t}\}_{t=1}^G$ и ненаблюдаемой: $X_m = \{X_{m_t}\}_{t=1}^G$, причем $X_t = X_{o_t} \cup X_{m_t}, \forall t \in \{1, \dots, G\}$. Предполагается, что 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 $t = \overline{1, G}$, $S = \{S_t\}_{t=1}^G$ — соответствует набору сигналов, испускаемых источниками в моменты времени $t = \overline{1, G}$, A — матрица управляющих векторов для равномерного линейного массива:

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

Сигналы, испускаемые источниками, также как и шумы на сенсорах, предполагаются стохастическими: $S_t \sim CN(\mathbf{O}_{M \times 1}, \mathbf{P})$, $t = \overline{1, G}$, $N_t \sim CN(\mathbf{O}_{L \times 1}, \mathbf{\Lambda})$. Матрицы \mathbf{P} 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 ;
- A_{o_t} — матрица, образованная теми строками матрицы A , которые соответствуют работающим сенсoram в момент времени t ;
- A_{m_t} — матрица, образованная теми строками матрицы A , которые соответствуют неисправным сенсoram в момент времени t ;
- $\mathbf{\Lambda}_{m_t}$ — ковариационная матрица шума на неисправных сенсорах в момент времени t ;
- $\mathbf{\Lambda}_{o_t}$ — ковариационная матрица шума на исправных сенсорах в момент времени t .

Составим ЕСМ-алгоритм (Expectation Conditional Maximization алгоритм) для двух случаев:

- Известный шум;
- Неизвестный шум.

§2 Известный шум

Воспользуемся ЕСМ-алгоритмом для того, чтобы определить значения параметров $\Psi = (\theta, \mathbf{P})$, пропущенные значения $X_m = \{X_{m_t}\}_{t=1}^G$ рассматриваются как латентные переменные. Наблюдения $X_t, t = \overline{1, \bar{L}}$ предполагаются независимыми и одинаково распределенными.

Инициализация параметров

Оценим вектор углов $\theta^{(0)}$ 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: $\theta_i^{(0)} = \nu + (i - 1) \cdot \frac{2\pi}{M}, i = \overline{1, M}$.

Диагональные элементы матрицы \mathbf{P} задаем с помощью равномерного распределения:

$$p_{jj} \sim \mathcal{U}([0.2; 5]) \quad (3)$$

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:

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

$$\begin{aligned} X_t &= AS_t + N_t \\ S_t &\sim CN(\mathbf{O}_{M \times 1}, \mathbf{P}) \\ X_t &\sim CN(\mathbf{O}_{L \times 1}, \mathbf{A} \mathbf{P} \mathbf{A}^H + \mathbf{\Lambda}) \\ X_t | S_t &\sim CN(AS_t, \mathbf{\Lambda}) \\ X_{o_t} &\sim CN(\mathbf{O}_{L \times 1}, A_{o_t} \mathbf{P} A_{o_t}^H + \mathbf{\Lambda}_{o_t}) \end{aligned}$$

$$P(S | \Psi) = \prod_{t=1}^G \frac{1}{\pi^M \text{Det}(\mathbf{\Lambda})} e^{-S_t^H (\mathbf{P})^{-1} S_t}, \quad (6)$$

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

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

$$P(X_o|\Psi) = \prod_{t=1}^G \frac{1}{\pi^{L_{o_t}} \text{Det}(\mathbf{\Lambda}_{o_t})} e^{-(X_{o_t})^H (\mathbf{\Lambda}_{o_t})^{-1} (X_{o_t})}, \quad (9)$$

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

$$\begin{cases} \mu_{X_{m_t}|x_{o_t}}^{(\tau)} = \hat{\Sigma}_{x_{m_t}, x_{o_t}}^{(\tau)} (\hat{\Sigma}_{x_{o_t}, x_{o_t}}^{(\tau)})^{-1} \cdot x_{o_t} \\ \Sigma_{x_{m_t}|x_{o_t}}^{(\tau)} = \hat{\Sigma}_{x_{m_t}, x_{m_t}}^{(\tau)} - \hat{\Sigma}_{x_{m_t}, x_{o_t}}^{(\tau)} (\hat{\Sigma}_{x_{o_t}, x_{o_t}}^{(\tau)})^{-1} \hat{\Sigma}_{x_{o_t}, x_{m_t}}^{(\tau)} \end{cases} \quad (10)$$

Оцениваем пропущенные значения для каждого наблюдения через условное математическое ожидание: $\tilde{x}_{m_t} = \mu_{x_{m_t}|x_{o_t}}^{(\tau)}$. $\tilde{x}_t^{(\tau)}$ — оценка x_t с учетом оценки пропусков. Параметры апостериорного распределения $P(S_t|X_{o_t} = x_{o_t}, X_{m_t} = \tilde{x}_{m_t}, \Psi)$ можно найти исходя из следующих формул

$$\begin{cases} \mu_{S_t|X_{o_t}=x_{o_t}, X_{m_t}=\tilde{x}_{m_t}, \Psi}^{(\tau)} = \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H \left(A^{(\tau-1)} \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H + \mathbf{\Lambda} \right)^{-1} \tilde{x}_t^{(\tau)} \\ \Sigma_{S_t|X_{o_t}=x_{o_t}, X_{m_t}=\tilde{x}_{m_t}, \Psi}^{(\tau)} = \mathbf{P}^{(\tau-1)} - \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H \left(A^{(\tau-1)} \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H + \mathbf{\Lambda} \right)^{-1} A^{(\tau-1)} \mathbf{P}^{(\tau-1)} \end{cases} \quad (11)$$

Рассчитаем оценку пространственной ковариационной матрицы (пользуемся результатами выкладок по детерминированной модели):

$$R^{(\tau)} = \frac{1}{G} \sum_{t=1}^G \left[\tilde{\Sigma}_t^{(\tau)} + \tilde{x}_t^{(\tau)} (\tilde{x}_t^{(\tau)})^H \right] \quad (12)$$

где $\tilde{\Sigma}_t^{(\tau)}$ — матрица размера $L \times L$, в которой все элементы являются нулями, за исключением тех, что стоят на пересечении строк с номерами $j_1 \in m_t$ и столбцов с номерами $j_2 \in m_t$: они заменены величиной $\Sigma_{x_{m_t}|x_{o_t}}^{(\tau)}$. 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.

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

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

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

Первый СМ-шаг

Оценим углы прибытия сигналов θ , но оставляем оценку ковариации сигналов \mathbf{P} фиксированной: $\mathbf{P} = \mathbf{P}^{(\tau-1)}$.

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

Тогда минимизируемая функция примет следующий вид:

$$\begin{aligned} \mathcal{J}(\theta) &= \sum_{t=1}^G \mathbb{E} \left[(X_t - A(\theta) S_t)^H \mathbf{\Lambda}^{-1} (X_t - A(\theta) S_t) | X_o = x_o \right] = \\ &= \sum_{t=1}^G \text{Tr} \left(\mathbf{\Lambda}^{-1} \mathbb{E} \left[(X_t - A(\theta) S_t)(X_t - A(\theta) S_t)^H | X_o = x_o \right] \right) = \\ &= \text{Tr} \left(\mathbf{\Lambda}^{-1} (R^{(\tau)} - \mathbf{A} \mathbf{P}^{\tau-1} A^H) \right) \\ \arg \min_{\theta} &= \text{Tr} \left(\mathbf{\Lambda}^{-1} (R^{(\tau)} - \mathbf{A} \mathbf{P}^{(\tau-1)} A^H) \right) \end{aligned} \tag{13}$$

Второй СМ-шаг

Оценим ковариацию сигналов \mathbf{P} , но оставляем оценку углов прибытия сигналов θ фиксированной: $\theta = \theta^{(\tau)}$

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

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

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

$X = \tilde{x}^{(\tau)}$ – оценка наблюдений, полученная с учетом Е-шага текущей итерации. Обозначим через M величину $\mathbb{E}[S_t S_t^H | X = \tilde{x}^{(\tau)}]$.

$$\frac{d}{d\mathbf{P}} \text{Tr}(\mathbf{P}^{-1} M) = -\mathbf{P}^{-1} M \mathbf{P}^{-1}$$

$$\frac{d\mathcal{K}(\mathbf{P})}{d\mathbf{P}} = -G\mathbf{P}^{-1} + \mathbf{P}^{-1}M\mathbf{P}^{-1}$$

Приравняем производную к нулю (функция по \mathbf{P} выпукла).

$$O = -G\mathbf{P}^{-1} + \mathbf{P}^{-1}M\mathbf{P}^{-1} \Rightarrow M = G\mathbf{P} \Rightarrow \mathbf{P}^{(\tau)} = \frac{1}{G} \sum_{t=1}^G \left(\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^H \right)$$

$$\mathbf{P}^{(\tau)} = \frac{1}{G} \sum_{t=1}^G \left(\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^H \right) \quad (14)$$

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

$$\mathbf{P}^{(\tau)} = \mathcal{D} \left[\frac{1}{G} \sum_{t=1}^G \left(\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^H \right) \right] \quad (15)$$

§3 Неизвестный шум

Е-шаг

М-шаг

Первый СМ-шаг

Второй СМ-шаг

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
2. Maximum Likelihood Estimation via the ECM Algorithm: A General Framework Xiao-Li Meng; Donald B. Rubin Biometrika, Vol. 80, No. 2 (Jun., 1993), 267-278.