

ЕМ-алгоритм, стохастическая модель сигнала

§1 Постановка проблемы

Предположим, у нас имеется линейная антенна решетка, состоящая из L сенсоров. Решетка принимает волны, направленные из K 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 $o_t \cup m_t = \{1, \dots, L\}$, $o_t \cap m_t = \emptyset, \forall t \in \{1, \dots, G\}$. 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:

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

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 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: $S_t \sim CN(\mathbf{0}_{K \times 1}, \boldsymbol{\Gamma})$, $t = \overline{1, G}$, $N_t \sim CN(\mathbf{0}_{L \times 1}, \boldsymbol{\Lambda})$. Матрицы $\boldsymbol{\Gamma}$ и $\boldsymbol{\Lambda}$ 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 $X_m = \{X_{m_t}\}_{t=1}^G$ рассматриваются как латентные переменные. Наблюдения $X_t, t = \overline{1, L}$ 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: $\theta_i^{(0)} = (\nu + (i-1) \cdot \frac{2\pi}{K}) \bmod 2\pi, i = \overline{1, K}$. 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 $j = \overline{1, K}$.

E-шаг

Требуется найти условное математическое ожидание с учетом текущей оценки параметров и апостериорного совместного распределения ненаблюденных/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:

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

$$\begin{aligned} X_t &= AS_t + N_t \\ S_t &\sim CN(\mathbf{0}_{M \times 1}, \Gamma) \\ X_t &\sim CN(\mathbf{0}_{L \times 1}, A\Gamma A^* + \Lambda) \\ X_t|S_t &\sim CN(AS_t, \Lambda) \\ X_{o_t} &\sim CN(\mathbf{0}_{L \times 1}, A_{o_t}\Gamma A_{o_t}^* + \Lambda_{o_t}) \end{aligned}$$

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

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

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

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

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

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

Пусть $\tilde{x}_t^{(\tau)}$ — вектор x_t , в котором пропущенные значения x_{m_t} оценены с помощью $\tilde{x}_{m_t}^{(\tau)}$, $\tilde{x}^{(\tau)}$ — реализация матрицы наблюдений x , в которой пропущенные значения x_{m_t} оценены с помощью $\tilde{x}_{m_t}^{(\tau)}$ для всех $t \in \overline{1, G}$.

Находим новую оценку ковариации наблюдений $\hat{\Sigma}_X^{(\tau+1)}$ как выборочную ковариацию восстановленного набора данных. 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

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

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

Заметим, что $\Sigma_{S_t|X_{o_t}=x_{o_t}, X_{m_t}=\tilde{x}_{m_t}^{(\tau)}}$ не зависит от величин, которые зависят от t , эта условная ковариация исходных сигналов вообще 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.

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

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

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

Оценим углы прибытия сигналов θ :

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

Тогда максимизируемая функция примет следующий вид:

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

Оценка УМО, полученная выше, была выведена в документе для детерминированной модели сигналов.

$$\theta^{(\tau)} = \arg \min_{\theta} \|\boldsymbol{\Lambda}^{-1/2}(\tilde{x}^{(\tau)} - A\tilde{S})\|_F^2 \quad (12)$$

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

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

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

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

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

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

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

Приравняем производную к нулю (функция по $\boldsymbol{\Gamma}$ выпукла).

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

$$\boldsymbol{\Gamma}^{(\tau)} = \frac{1}{G} \sum_{t=1}^G \left(\boldsymbol{\Sigma}_{S_t|X_t} + \boldsymbol{\mu}_{S_t|X_t} (\boldsymbol{\mu}_{S_t|X_t})^* \right) \quad (13)$$

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

$$\boldsymbol{\Gamma}^{(\tau)} = \frac{1}{G} M_{S|X} \cdot M_{S|X}^* + \boldsymbol{\Sigma}_{S_1|X_1}, \quad (14)$$

где $M_{S|X}$ – матрица, составленная из столбцов $\boldsymbol{\mu}_{S_t|X_t}, t = \overline{1, G}$. Предполагая, что сигналы некоррелированы, изменим оценку так, чтобы учесть лишь диагональные элементы:

$$\boldsymbol{\Gamma}^{(\tau)} = \mathcal{D} \left[\frac{1}{G} M_{S|X} \cdot M_{S|X}^* + \boldsymbol{\Sigma}_{S_t|X_t} \right] \quad (15)$$

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