ECM-алгоритм, стохастическая модель сигнала

§1 Постановка проблемы

Предположим, у нас имеется линейная антенная 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 $o_t \cup m_t = \{1, ..., L\}, o_t \cap m_t = \varnothing, \forall t \in \{1, ..., G\}$. 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:

$$X = AS + N, \tag{1}$$

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:

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

Сигналы, испускаемые источниками, также как и шумы на сенсорах, предполагаются стохастическими: $S_t \sim CN(\mathbf{O}_{K\times 1}, \mathbf{P}), t = \overline{1, G}$, $N_t \sim CN(\mathbf{O}_{L\times 1}, \mathbf{\Lambda})$. 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 сенсорам в момент времени t;
- A_{m_t} — матрица, образованная теми строками матрицы A, которые соответствуют неисправным сенсорам 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 ECM-алгоритм (Expectation Conditional Maximization алгоритм) для двух случаев:

- Известный шум;
- Неизвестный шум.

§2 Известный шум

Воспользуемся ECM-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, $t = \overline{1, L}$ предполагаются независимыми и одинаково распределенными.

Инициализация параметров

Оценим вектор углов $\theta^{(0)}$ следующим образом:

1. Выберем число ν, которое будет соответствовать первому компоненту вектора $\theta^{(0)}$:
$$\nu \sim \mathcal{U}([-\pi; \pi]); \tag{2}$$
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 $\mathbf{P}$ задаем с помощью равномерного распределения:
$$p_{jj} \sim \mathcal{U}([0.2; 5]) \tag{3}$$
где $j = \overline{1, K}$.

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:
$$P((X_m, S)|X_o = x_o, \Psi) = P(X_m|X_o = x_o, \Psi) \cdot P(S|X_o = x_o, X_m = \widetilde{x}_m, \Psi) \tag{5}$$

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

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

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

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

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

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

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

Пусть $\widetilde{x}^{(\tau)}_t$ — вектор x_t, в котором пропущенные значения x_{m_t} оценены с помощью $\widetilde{x}^{(\tau)}_{m_t}$, $\widetilde{x}^{(\tau)}$ — реализация матрицы наблюдений x, в которой пропущенные значения x_{m_t} оценены с помощью $\widetilde{x}^{(\tau)}_{m_t}$ для всех $t \in \overline{1, G}$.
Находим новую оценку ковариации наблюдений $\hat{\Sigma}^{(\tau+1)}_X$ как выборочную ковариацию восстановленного набора данных. Параметры апостериорного распределения $P(S_t|X_{o_t} = x_{o_t}, X_{m_t} = \widetilde{x}_{m_t}, \Psi)$ можно найти исходя из следующих формул

$$\begin{cases} \mu^{(\tau)}_{S_t|X_{o_t}=x_{o_t}, X^{(\tau)}_{m_t}=\widetilde{x}_{m_t}, \Psi} = \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H \Big(A^{(\tau-1)} \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H + \mathbf{\Lambda} \Big)^{-1} \widetilde{x}^{(\tau)}_t \\ \Sigma^{(\tau)}_{S_t|X_{o_t}=x_{o_t}, X^{(\tau)}_{m_t}=\widetilde{x}_{m_t}, \Psi} = \mathbf{P}^{(\tau-1)} - \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H \Big(A^{(\tau-1)} \mathbf{P}^{(\tau-1)} (A^{(\tau-1)})^H + \mathbf{\Lambda} \Big)^{-1} A^{(\tau-1)} \mathbf{P}^{(\tau-1)} \end{cases} \tag{11}$$

Заметим, что $\Sigma_{S_t|X_{o_t}=x_{o_t}, X^{(\tau)}_{m_t}}$ 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.

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

М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

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

Первый CM-шаг

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

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

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

Тогда максимизируемая функция примет следующий вид:

$$\mathcal{J}(\theta) = \sum_{t=1}^{G}\mathbb{E}\Big[(X_t - A(\theta)S_t)^H\mathbf{\Lambda}^{-1}(X_t - A(\theta)S_t)|X_o = x_o\Big] =$$
$$= -||\mathbf{\Lambda}^{-1/2}(\widetilde{x}^{(\tau)} - AS)||_F^2.$$

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

$$\theta^{(\tau)} = \arg\min_{\theta} ||\mathbf{\Lambda}^{-1/2}(\widetilde{x}^{(\tau)} - AS)||_F^2 \tag{12}$$

Второй CM-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 $\log(S|\Psi)$.

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

$$\frac{d}{d\mathbf{P}}\log(\mathrm{Det}(\mathbf{P})) = \mathbf{P}^{-1}$$

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

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

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

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

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

$$\mathbf{P}^{(\tau)} = \frac{1}{G}\sum_{t=1}^{G}\left(\Sigma_{S_t|X_t} + \mu_{S_t|X_t}(\mu_{S_t|X_t})^H\right) \tag{13}$$

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

$$\mathbf{P}^{(\tau)} = \frac{1}{G}M_{S|X} \cdot M_{S|X}^H + \Sigma_{S_1|X_1}, \tag{14}$$

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

$$\mathbf{P}^{(\tau)} = \mathcal{D}\left[\frac{1}{G}M_{S|X} \cdot M_{S|X}^H + \Sigma_{S_1|X_1}\right] \tag{15}$$

.

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

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.