

# ЕМ-алгоритм, стохастическая модель сигнала

## Постановка задачи

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)  $\theta$ , практически не изменяющиеся во времени. По итогам измерений было получено  $T$  снимков полученного сигнала, причем ввиду стохастических технических сбоев, связанных с сенсорами, большая часть таких снимков содержит помимо надежных данных ненадежные, которые в рамках данной задачи рассматриваются как пропуски. Пусть  $X$  — набор наблюдений, полученных сенсорами в моменты времени  $t = 1, \dots, T$ ,  $X_t$  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 набор наблюдений  $X$  состоит из доступной части  $X_o = \{X_{t,o_t}\}_{t=1}^T$  и недоступной:  $X_m = \{X_{t,m_t}\}_{t=1}^T$ , причем  $o_t \cup m_t = \{1, \dots, T\}$ ,  $o_t \cap m_t = \emptyset, \forall t \in \{1, \dots, T\}$ . Предполагается, что  $\nexists o_t : o_t = \emptyset$ , т.е. нет таких наблюдений, которые состоят лишь из недоступной части. Набор  $X$  является результатом следующей модели наблюдений:

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

где

$$X = [X_1, \dots, X_T], S = [S_1, \dots, S_T], N = [N_1, \dots, N_T], \quad (2)$$

и  $X_t \in \mathbb{C}^{L \times T}$ ,  $S_t \in \mathbb{C}^{K \times T}$ ,  $N_t \in \mathbb{C}^{L \times T}$  — векторы-столбцы, соответствующие наблюдениям, источникам и шумам в момент времени  $t = 1, \dots, T$ ,  $A$  — матрица векторов направленности для равномерного линейного массива:

$$A(\theta) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-2j\pi \frac{d}{\lambda} \sin(\theta_1)} & e^{-2j\pi \frac{d}{\lambda} \sin(\theta_2)} & \dots & e^{-2j\pi \frac{d}{\lambda} \sin(\theta_K)} \\ \dots & \dots & \ddots & \vdots \\ e^{-2j\pi(L-1) \frac{d}{\lambda} \sin(\theta_1)} & e^{-2j\pi(L-1) \frac{d}{\lambda} \sin(\theta_2)} & \dots & e^{-2j\pi(L-1) \frac{d}{\lambda} \sin(\theta_K)} \end{bmatrix}. \quad (3)$$

Сигналы, испускаемые источниками, шумы на сенсорах и наблюдения предполагаются стохастическими:  $S_t \sim \mathcal{CN}(\mathbf{O}_{K \times 1}, \mathbf{\Gamma})$ ,  $N_t \sim \mathcal{CN}(\mathbf{O}_{L \times 1}, \mathbf{\Lambda})$ ,  $X_t \sim \mathcal{CN}(\mathbf{O}_{L \times 1}, \mathbf{A}\mathbf{A}^* + \mathbf{\Lambda})$ ,  $t = 1, \dots, T$ . Матрицы  $\mathbf{\Gamma}$  и  $\mathbf{\Lambda}$  предполагаются диагональными, т.е. сигналы не коррелированы между собой, шумы также не коррелированы между собой, корреляция между сигналами и шумами также отсутствует. Сигналы предполагаются узкополосными. Составим ЕМ-алгоритм для случая известного шума.

## ЕМ-алгоритм для известного шума

Воспользуемся ЕМ-алгоритмом для того, чтобы определить значения параметров  $\Upsilon = (\theta, \Gamma)$ , пропущенные значения  $X_m = \{X_{t,m_t}\}_{t=1}^T$  и сигналы  $S$  рассматриваются как латентные переменные. Наблюдения  $X_t$ ,  $t = 1, \dots, T$  предполагаются независимыми и одинаково распределенными.

### Е-шаг

Требуется найти математическое ожидание полного правдоподобия с учетом текущей оценки параметров и апостериорного совместного распределения пропущенных значений в наблюдениях  $X_m$  и сигналов  $S$

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

Преобразуем выражение, учитывая тот факт, что наблюдения являются независимыми, и обозначив через  $\mathcal{I}$  условную информацию  $X_o = x_o$ ,  $\Upsilon^{(\tau-1)}$  и через  $\mathcal{I}_t$  условную информацию  $X_{t,o_t} = x_{t,o_t}$ ,  $\Upsilon^{(\tau-1)}$ :

$$\begin{aligned} & \mathbb{E}_{(X_m, S)|\mathcal{I}}[\log P(X, S)] = \\ & \mathbb{E}_{(X_m, S)|\mathcal{I}}[\log P(X|S) + \log P(S)] = \\ & \left[ \sum_{t=1}^T \mathbb{E}_{(X_{t,m_t}, S_t)|\mathcal{I}_t} \log[P(X_t|S_t)] + \sum_{t=1}^T \mathbb{E}_{(X_{t,m_t}, S_t)|\mathcal{I}_t} [\log P(S_t)] \right] = \\ & -T \left[ \log |\Lambda| + \text{Tr}(\Lambda^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[X X^*]) - 2 \text{Tr}(\Lambda^{-1} A(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[X S^*]) \right. \\ & \left. + \text{Tr}(\Lambda^{-1} A(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[S S^*] A^*(\theta)) + \log |\Gamma| + \text{Tr}(\Gamma^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[S S^*]) \right]. \end{aligned}$$

Выражение  $\mathbb{E}_{(X_{t,m_t}, S_t)|\mathcal{I}_t}[\cdot]$  соответствуют случаю максимально полного набора латентных переменных: некоторые наблюдения в наборе могут не содержать пропуски ( $X_{t,m_t} = \emptyset$ ), и тогда это выражение упрощается до  $\mathbb{E}_{S_t|\mathcal{I}_t}[\cdot]$ . Для нахождения условных моментов, указанных в формуле выше, требуется найти апостериорное распределение скрытых переменных.

### Апостериорное распределение латентных переменных для неполных наблюдений

Воспользуемся формулой произведения плотностей:

$$P(X_{t,m_t}, S_t|\mathcal{I}_t) = P(X_{t,m_t}|\mathcal{I}_t) \cdot P(S_t|X_{t,m_t}, \mathcal{I}_t). \quad (5)$$

Сначала найдем апостериорное распределение  $P(X_m|\mathcal{I})$ . Для достижения этой цели, для каждой пары  $\{(o_t, m_t) : m_t \neq \emptyset\}$  создадим разбиение оценки ковариационной матрицы наблюдений  $\hat{R}$  на блоки, индуцированное этим разбиением множества индексов, оно имеет следующий вид:

$$\hat{R} = \begin{pmatrix} \hat{R}_{o_t, o_t} & \hat{R}_{o_t, m_t} \\ \hat{R}_{m_t, o_t} & \hat{R}_{m_t, m_t} \end{pmatrix}, \quad (6)$$

где каждый блок определяется как

$$\hat{R}_{a,b} = (\hat{R}_{ij})_{i \in a, j \in b}.$$

Для каждого наблюдения, содержащего пропуски, требуется найти апостериорное распределение пропущенных значений,  $P(X_{t,m_t}|X_{t,o_t} = x_{t,o_t}, \Upsilon^{(\tau-1)})$ ,  $t = 1, \dots, T$ ,  $m_t \neq \emptyset$ . Обозначим через  $\mathcal{I}_t$  условную информацию  $X_{t,o_t} = x_{t,o_t}$ ,  $\Upsilon^{(\tau-1)}$ . Параметры апостериорного распределения  $P(X_{t,m_t}|\mathcal{I}_t)$ ,  $t = 1, \dots, T$  на итерации  $\tau$  можно найти следующим образом:

$$\begin{cases} \mathbb{E}[X_{t,m_t}|\mathcal{I}_t] = \hat{R}_{m_t, o_t} \left( \hat{R}_{o_t, o_t} \right)^{-1} \cdot x_{t,o_t}, \\ \text{Cov}(X_{t,m_t}|\mathcal{I}_t) = \hat{R}_{m_t, m_t} - \hat{R}_{m_t, o_t} \left( \hat{R}_{o_t, o_t} \right)^{-1} \hat{R}_{o_t, m_t}, \end{cases} \quad (7)$$

где  $\hat{\mathbf{R}}_{o_t, o_t} = \hat{\mathbf{R}}_{o_t, o_t}^{(\tau-1)}$ ,  $\hat{\mathbf{R}}_{o_t, m_t} = \hat{\mathbf{R}}_{o_t, m_t}^{(\tau-1)}$ ,  $\hat{\mathbf{R}}_{m_t, o_t} = \hat{\mathbf{R}}_{m_t, o_t}^{(\tau-1)}$ ,  $\hat{\mathbf{R}}_{m_t, m_t} = \hat{\mathbf{R}}_{m_t, m_t}^{(\tau-1)}$ .

Для каждого наблюдения  $X_t$ , содержащего пропуски, оценим условный второй начальный момент  $\mathbb{E}[X_t X_t^* | \mathcal{I}_t]$ :

$$\begin{aligned} \mathbb{E}[X_t X_t^* | \mathcal{I}_t] &= \mathbb{E}[X_t | \mathcal{I}_t] \cdot \mathbb{E}[X_t^* | \mathcal{I}_t] + \text{Cov}(X_t | \mathcal{I}_t) \\ &= \mathbb{E}[X_t | \mathcal{I}_t] \cdot \mathbb{E}[X_t^* | \mathcal{I}_t] + \begin{pmatrix} \mathbf{O}_{o_t, o_t} & \mathbf{O}_{o_t, m_t} \\ \mathbf{O}_{m_t, o_t} & \text{Cov}(X_{t, m_t} | \mathcal{I}_t) \end{pmatrix}, \end{aligned}$$

где  $\mathbf{O}_{o_t, o_t}$ ,  $\mathbf{O}_{o_t, m_t}$ ,  $\mathbf{O}_{m_t, o_t}$  — нулевые блочные матрицы. Разбиение указанной матрицы на четыре блока, три из которых состоят из нулей, индуцировано разбиением множества индексов  $\{1, \dots, L\}$  на множества  $o_t, m_t$ .

Нам недоступны значения  $X_{t, m_t}$ , ввиду этого мы не можем найти  $P(S_t | \mathcal{I}_t, X_{t, m_t})$ , однако ввиду того, что модель линейная гауссовская, мы можем найти параметры распределения  $P(S_t | \mathcal{I}_t)$ ,  $t = 1, \dots, T$ . Параметры апостериорного распределения  $P(S_t | \mathcal{I}_t)$ ,  $t = 1, \dots, T$  можно найти исходя из следующих формул, если  $m_t \neq \emptyset$ :

$$\begin{cases} \mathbb{E}[S_t | \mathcal{I}_t] = \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbb{E}[X_t | \mathcal{I}_t], \\ \text{Cov}(S_t | \mathcal{I}_t) = \mathbf{\Gamma} - \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma} + \mathbf{\Gamma} \mathbf{A}^* \hat{\mathbf{R}}^{-1} \text{Cov}(X_t | \mathcal{I}_t) \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}, \end{cases} \quad (8)$$

где  $\mathbf{A} = \mathbf{A}(\theta^{(\tau-1)})$ ,  $\mathbf{\Gamma} = \mathbf{\Gamma}^{(\tau-1)}$ ,  $\hat{\mathbf{R}} = \hat{\mathbf{R}}^{(\tau-1)}$ . Для вывода этих формул используются свойство башни условного математического ожидания (tower property) и закон полной дисперсии (подробности приведены в приложении 2).

Оценим  $\mathbb{E}_{(X_{t, m_t}, S_t) | \mathcal{I}_t} [S_t S_t^*]$ :

$$\mathbb{E}_{(X_{t, m_t}, S_t) | \mathcal{I}_t} [S_t S_t^*] = \mathbb{E}[S_t | \mathcal{I}_t, X_{t, m_t}] \cdot \mathbb{E}[S_t^* | \mathcal{I}_t, X_{t, m_t}] + \text{Cov}(S_t | \mathcal{I}_t, X_{t, m_t}), \quad (9)$$

Оценим  $\mathbb{E}_{(X_{t, m_t}, S_t) | \mathcal{I}_t} [X_t S_t^*]$ :

$$\mathbb{E}_{(X_{t, m_t}, S_t) | \mathcal{I}_t} [X_t S_t^*] = \mathbb{E}[X_t X_t^* | \mathcal{I}_t] \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}, \quad (10)$$

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

## Апостериорное распределение латентных переменных для полных наблюдений

Если  $m_t = \emptyset$ , то апостериорное распределение упрощается следующим образом:  $P(X_{t, m_t}, S_t | \mathcal{I}_t) = P(S_t | \mathcal{I}_t)$ . Параметры апостериорного распределения  $P(S_t | \mathcal{I}_t)$ ,  $t = 1, \dots, T$ :

$$\begin{cases} \mathbb{E}[S_t | \mathcal{I}_t] = \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbb{E}[X_t | \mathcal{I}_t], \\ \text{Cov}(S_t | \mathcal{I}_t) = \mathbf{\Gamma} - \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}. \end{cases} \quad (11)$$

При этом:

$$\mathbb{E}_{S_t | \mathcal{I}_t} [X_t X_t^*] = \mathbb{E}[X_t | \mathcal{I}_t] \cdot \mathbb{E}[X_t^* | \mathcal{I}_t] = x_t x_t^*, \quad (12)$$

$$\mathbb{E}_{S_t | \mathcal{I}_t} [X_t S_t^*] = \mathbb{E}[X_t | \mathcal{I}_t] \cdot \mathbb{E}[S_t^* | \mathcal{I}_t] = x_t \cdot \mathbb{E}[S_t^* | \mathcal{I}_t] = x_t \cdot x_t^* \cdot \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}, \quad (13)$$

$$\mathbb{E}_{S_t | \mathcal{I}_t} [S_t S_t^*] = \mathbb{E}[S_t | \mathcal{I}_t] \cdot \mathbb{E}[S_t^* | \mathcal{I}_t] + \text{Cov}(S_t | \mathcal{I}_t). \quad (14)$$

## Агрегация

Теперь оценим вторые начальные моменты  $\mathbb{E}_{(X_m, S)|\mathcal{I}}[XX^*]$ ,  $\mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]$ ,  $\mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*]$ :

$$\mathbb{E}_{(X_m, S)|\mathcal{I}}[XX^*] = \frac{1}{T} \sum_{t=1}^T \mathbb{E}[X_t X_t^* | \mathcal{I}_t], \quad (15)$$

$$\mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*] = \frac{1}{T} \left[ \sum_{t \in T_1} \mathbb{E}_{(X_t, m_t, S_t)|\mathcal{I}_t}[S_t S_t^*] + \sum_{t \in T_2} \mathbb{E}_{S_t|\mathcal{I}_t}[S_t S_t^*] \right], \quad (16)$$

$$\mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*] = \frac{1}{T} \left[ \sum_{t \in T_1} \mathbb{E}_{(X_t, m_t, S_t)|\mathcal{I}_t}[X_t S_t^*] + \sum_{t \in T_2} \mathbb{E}_{S_t|\mathcal{I}_t}[X_t S_t^*] \right], \quad (17)$$

где  $T_1 = \{y \in \mathbb{N} : m_y \neq \emptyset, 1 \leq y \leq T\}$ ,  $T_2 = \{y \in \mathbb{N} : m_y = \emptyset, 1 \leq y \leq T\}$ .

## М-шаг

Требуется найти наилучшую оценку параметров, решив следующую задачу оптимизации:

$$\begin{aligned} \Upsilon^{(\tau)} &= \arg \max_{\Upsilon} \mathbb{E}_{(X_m, S)|\mathcal{I}}[\log P(X, S)] = \\ &= \arg \min_{\Upsilon} T \left[ \log |\mathbf{\Lambda}| + \text{Tr}(\mathbf{\Lambda}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[XX^*]) - 2 \text{Tr}(\mathbf{\Lambda}^{-1} \mathbf{A}(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*]) \right. \\ &\quad \left. + \text{Tr}(\mathbf{\Lambda}^{-1} \mathbf{A}(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \mathbf{A}^*(\theta)) + \log |\mathbf{\Gamma}| + \text{Tr}(\mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) \right]. \end{aligned}$$

Оценим угловые координаты источников  $\theta$ :

$$\begin{aligned} \theta^{(\tau)} &= \arg \min_{\theta} \mathcal{Q}_1(\theta | \theta^{(\tau-1)}) = \arg \min_{\theta} \left[ -2 \text{Tr}(\mathbf{\Lambda}^{-1} \mathbf{A}(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*]) \right. \\ &\quad \left. + \text{Tr}(\mathbf{\Lambda}^{-1} \mathbf{A}(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \mathbf{A}^*(\theta)) \right] = \\ &= \arg \min_{\theta} \|\mathbf{\Lambda}^{-1/2} (\mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*] - \mathbf{A}(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*])\|_F^2. \end{aligned}$$

Эту задачу можно решить численно, подробности приведены в приложении 3.

Оценим ковариацию сигналов  $\mathbf{\Gamma}$  с учетом предположения о некоррелированности сигналов:

$$\mathbf{\Gamma}^{(\tau)} = \arg \min_{\mathbf{\Gamma}} \mathcal{Q}_2(\mathbf{\Gamma} | \mathbf{\Gamma}^{(\tau-1)}) = T \left[ \log |\mathbf{\Gamma}| + \text{Tr}(\mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) \right] = \mathcal{D}[\mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]].$$

Обновляем оценку ковариации наблюдений с учетом полученных оценок параметров:

$$\hat{\mathbf{R}}^{(\tau)} = \mathbf{A}(\theta^{(\tau)}) \mathbf{\Gamma}^{(\tau)} \mathbf{A}^*(\theta^{(\tau)}) + \mathbf{\Lambda}. \quad (18)$$

# Результаты численного эксперимента

## Первые результаты, предварительные

### Первый набор начальных условий

$L = 25, K = 1, T = 12, 5$  сенсоров неисправны, 50% пропусков для каждого из них,  
 $\Gamma = 0.5 \cdot E_K, \Lambda = 8.1 \cdot E_L, \theta = [0.7] \approx [40.107^\circ]$ .

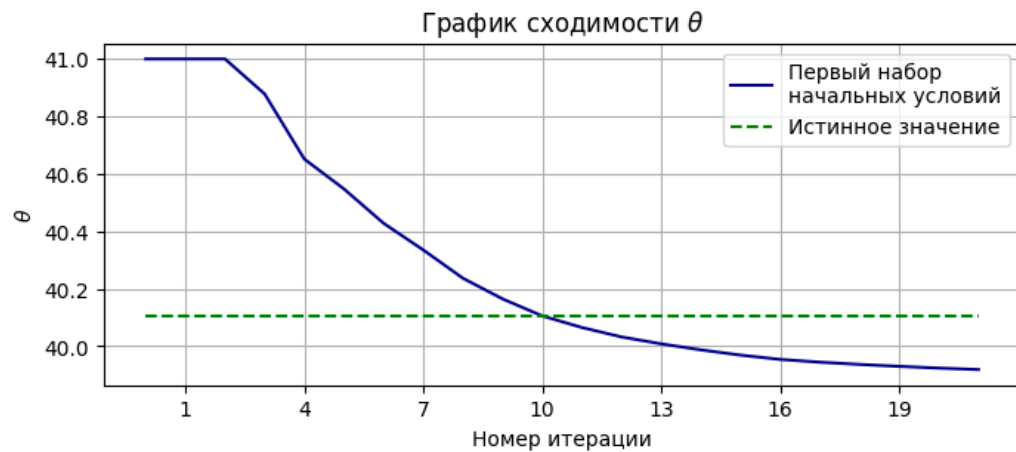


Рис. 1: График сходимости угловой координаты для первого набора начальных условий

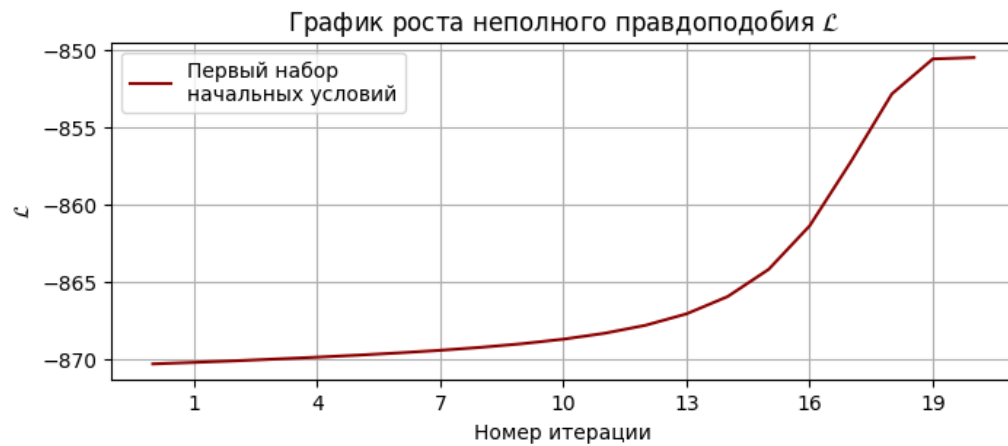


Рис. 2: График роста  $\log(p(X_o|\Psi))$  для первого набора начальных условий

### Второй набор начальных условий

$L = 25, K = 1, T = 11$ , 8 сенсоров неисправны, 50% пропусков для каждого из них,  
 $\Gamma = 0.5 \cdot E_K, \Lambda = 6.1 \cdot E_L, \theta = [0.7] \approx [40.107^\circ]$ .

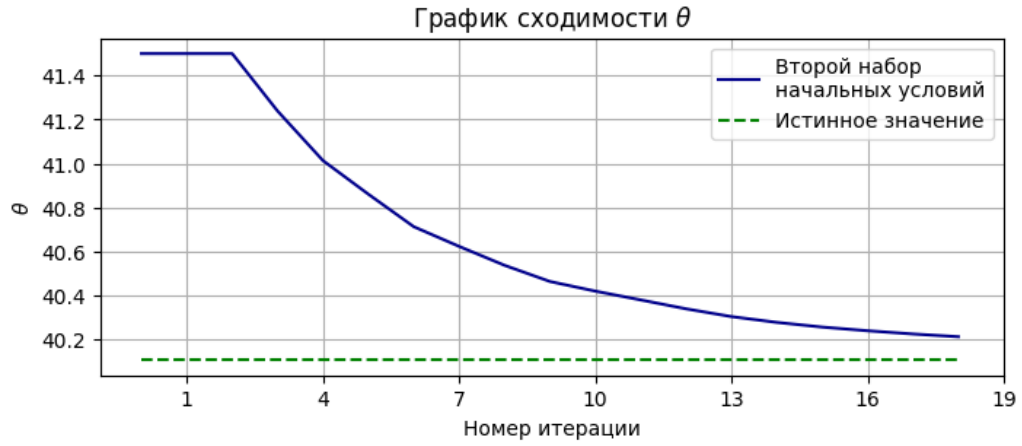


Рис. 3: График сходимости угловой координаты для второго набора начальных условий

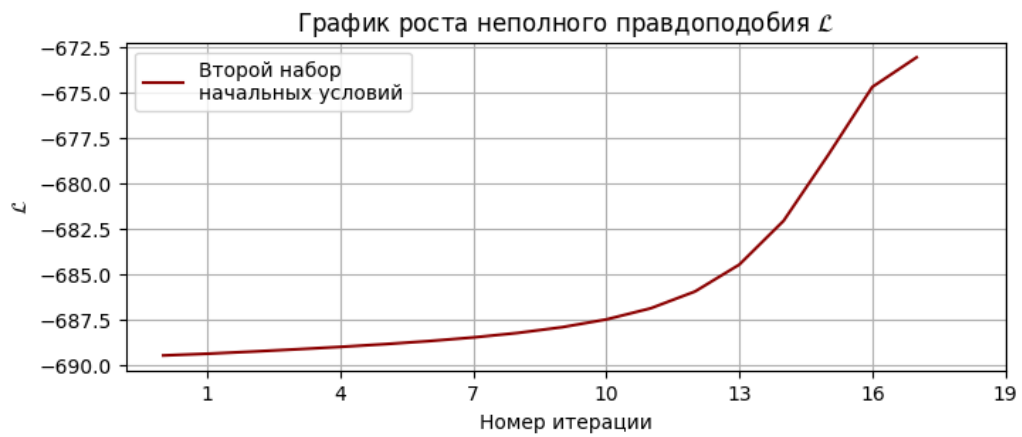


Рис. 4: График роста  $\log(p(X_o|\Psi))$  для второго набора начальных условий

## СПИСОК ИСТОЧНИКОВ

1. Dempster A. P., Laird N. M., Rubin D. B. Maximum likelihood from incomplete data via the EM algorithm // Journal of the Royal Statistical Society: Series B (Methodological). – 1977. – Vol. 39. – No. 1. – P. 1–38.
2. Wu C. F. J. On the convergence properties of the EM algorithm // The Annals of Statistics. – 1983. – Vol. 11. – No. 1. – P. 95–103.
3. Louis T. A. Finding the observed information matrix when using the EM algorithm // Journal of the Royal Statistical Society: Series B (Methodological). – 1982. – Vol. 44. – No. 2. – P. 226–233.
4. Ross S. M. A first course in probability. 8th ed. – Upper Saddle River, N. J.: Prentice Hall, 2010. – 640p., p. 348.
5. Little R. J. A., Rubin D. B. Statistical analysis with missing data 2nd ed. – Hoboken, N. J.: Wiley-Interscience, Jon Wiley & Sons, Inc., 2002. – 381p.
6. Schreier P. J., Scharf L. L. Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals – Cambridge: Cambridge University Press, 2010. – 309p., pp. 41–42.
7. Nocedal J., Wright S.J. Numerical Optimization. 2nd ed. – New York: Springer, 2006. – 664p., pp. 10–37.

## Приложение №1. Инициализация

Оценим ковариационную матрицу  $\hat{\mathbf{R}}^{(0)}$  следующим образом: если  $\sum_t \mathbf{I}_{x_t=x_{t,o_t}} \geq L$ , оцениваем  $\hat{\mathbf{R}}^{(0)}$  по выборке, образованной  $x_t : x_t = x_{t,o_t}$ . В противном случае используется следующий подход: пусть  $x^l$  – строка матрицы  $X$ , соответствующая сенсору  $l, l = 1, \dots, L$ ,  $x_o^l$  – выборка, образованная доступными наблюдениями в строке  $l$ ,  $x_m^l$  – выборка, образованная недоступными наблюдениями в строке  $l$ ,  $x_t^l$  – компонента наблюдения  $x_t$ , соответствующая сенсору  $l$ . Для всех недоступных значений построим следующую оценку: если  $x_t^l \in x_m^l, l = 1, \dots, L, t = 1, \dots, T$ , то  $\hat{x}_t^l = \bar{x}_o^l$ . Оцениваем ковариацию по такому оцененному набору  $\hat{x}$ :

$$\hat{\mathbf{R}}^{(0)} = \text{Cov}(\hat{x}). \quad (19)$$

Оценим вектор угловых координат источников  $\theta^{(0)}$  следующим образом:

1. Выберем число  $\nu$ , которое будет соответствовать первому компоненту вектора  $\theta^{(0)}$ :

$$\nu \sim \mathcal{U}([- \pi; \pi]); \quad (20)$$

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

Альтернативный подход: используем  $\hat{\mathbf{R}}^{(0)}$  для оценки  $\theta^{(0)}$  с помощью алгоритма MUSIC, используем эту оценку на первой итерации мултистарта ЕМ, на следующих используем оценки в окрестности указанной.

Начальную ковариацию сигналов  $\mathbf{\Gamma}^{(0)}$  получаем на основе метода наименьших квадратов с помощью  $\theta^{(0)}, \hat{\mathbf{R}}^{(0)}$ . Матрица  $A(\theta^{(0)})$  – матрица векторов направленности и представима в следующем виде:

$$A(\theta^{(0)}) = [a(\theta_1^{(0)}), \dots, a(\theta_K^{(0)})], \quad (21)$$

где  $a(\theta_k^{(0)})$  – вектор направленности для источника  $k$ . Перед применением МНК каждый такой вектор отнормируем:

$$\check{a}(\theta_k^{(0)}) = \frac{a(\theta_k^{(0)})}{\|a(\theta_k^{(0)})\|}, k = 1, \dots, K, \quad (22)$$

сформируем из них нормированную матрицу векторов направленности:

$$\check{A}(\theta^{(0)}) = [\check{a}(\theta_1^{(0)}), \dots, \check{a}(\theta_K^{(0)})], \quad (23)$$

применяем МНК для получения нормированной ковариации мощностей:

$$\check{\mathbf{\Gamma}}^{(0)} = \mathcal{D}[\check{A}(\theta^{(0)})^+ (\hat{\mathbf{R}}^{(0)} - \mathbf{\Lambda})(\check{A}(\theta^{(0)})^+)^*], \quad (24)$$

где  $\mathcal{D}[\cdot]$  – диагональное приближение матрицы,  $(\cdot)^+$  – псевдообратная матрица. Такой подход не гарантирует, что все элементы матрицы на главной диагонали будут неотрицательными, поэтому используем следующую корректировку:  $\check{\mathbf{\Gamma}}_l^{(0)} = \max(\check{\mathbf{\Gamma}}_l^{(0)}, \varepsilon), l = 1, \dots, L$ . Переходим к ненормированной мощности:

$$\mathbf{\Gamma}_l^{(0)} = \frac{\check{\mathbf{\Gamma}}_l^{(0)}}{\|a(\theta_k^{(0)})\|_l}, l = 1, \dots, L. \quad (25)$$

Требуется гарантировать, что при выбранной инициализации на Е-шаге не будет получена условная ковариация сигналов, не являющаяся положительной определенной. Вычисляем оценку условной ковариации сигналов, используя начальную оценку, и проверяем, выполняется ли свойство неотрицательной определенности:

$$\mathbf{\Gamma}^{(0)} - (\mathbf{\Gamma}^{(0)})^* A(\theta^{(0)})^* (\hat{\mathbf{R}}^{(0)})^{-1} A(\theta^{(0)}) \mathbf{\Gamma}^{(0)} \succeq 0, \quad (26)$$



пока это свойство не выполняется, домножаем матрицу на 0.5:  $\mathbf{\Gamma}^{(0)} = 0.5 \cdot \mathbf{\Gamma}^{(0)}$  и повторно проверяем выполненность вышеуказанного свойства.

## Приложение №2. Об апостериорном распределении сигналов, полученном на Е-шаге

В рамках Е-шага необходимо вычислить  $\mathbb{E}[S_t|X_{t,o_t} = x_{t,o_t}]$ ,  $\text{Cov}(S_t|X_{t,o_t} = x_{t,o_t})$ . Вычисление  $\mathbb{E}[S_t|X_{t,o_t} = x_{t,o_t}]$  реализуется так:

$$\begin{aligned}\mathbb{E}[S_t|X_{t,o_t} = x_{t,o_t}] &= \mathbb{E}[\mathbb{E}[S_t|X_t]|X_{t,o_t} = x_{t,o_t}] \\ &= \mathbb{E}[\mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} X_t|X_{t,o_t} = x_{t,o_t}] = \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbb{E}[X_t|X_{t,o_t} = x_{t,o_t}].\end{aligned}$$

Для нахождения  $\text{Cov}(S_t|X_{t,o_t} = x_{t,o_t})$  может быть использовано равенство, называемое законом полной дисперсии (Law of total variance, может быть найдено в Sheldon Russ, A first course in probability, 8th ed., 2010):

$$\text{Cov}(W) = \mathbb{E}[\text{Cov}(W|Y)] + \text{Cov}(\mathbb{E}[W|Y]). \quad (27)$$

Пусть  $\mathcal{F}_Z = \sigma(Z)$ : – сигма-алгебра, порожденная  $Z$ . Тогда все числовые характеристики  $W$ , включая условные моменты и ковариации относительно  $Y$ , можно обусловить на  $\mathcal{F}_Z$ . И, соответственно, из закона полной ковариации следует:

$$\text{Cov}(W|\mathcal{F}_Z) = \mathbb{E}[\text{Cov}(W|Y, Z)|\mathcal{F}_Z] + \text{Cov}(\mathbb{E}[W|Y, Z]|\mathcal{F}_Z). \quad (28)$$

Вектора  $W$ ,  $Y$ ,  $Z$  предполагаются комплексными гауссовскими (случай круговой симметрии), зависимость между  $Z$  и  $Y$  линейная,  $\sigma(Z) \subseteq \sigma(Y)$ , предполагается, что  $Z = CY$ , где  $C$  – булев селектор. Для линейной гауссовской модели знание  $Z$  не уменьшает условную ковариацию  $W|Y$  и не влияет на условное математическое ожидание  $\mathbb{E}[W|Y, Z]$ , поскольку  $Z$  не содержит никакой новой информации, которой не было бы в  $Y$ :

$$\begin{aligned}\text{Cov}(W|Y, Z) &= \text{Cov}(W|Y), \\ \mathbb{E}[W|Y, Z] &= \mathbb{E}[W|Y].\end{aligned}$$

Получаем:

$$\text{Cov}(W|\mathcal{F}_Z) = \mathbb{E}[\text{Cov}(W|Y)|\mathcal{F}_Z] + \text{Cov}(\mathbb{E}[W|Y]|\mathcal{F}_Z). \quad (29)$$

Если  $Z = z$ , имеем:

$$\text{Cov}(W|Z = z) = \mathbb{E}[\text{Cov}(W|Y)|Z = z] + \text{Cov}(\mathbb{E}[W|Y]|Z = z). \quad (30)$$

В рамках исходной задачи по оцениванию DoA и ковариации сигналов,  $W$  соответствует величине  $S_t$ ,  $Y$  соответствует величине  $X_t$ ,  $Z$  соответствует величине  $X_{t,o_t}$ ,  $z$  соответствует величине  $x_{t,o_t}$ ,  $t = 1, \dots, T$ .

Получаем:

$$\text{Cov}(S_t|X_{t,o_t} = x_{t,o_t}) = \mathbb{E}[\text{Cov}(S_t|X_t)|X_{t,o_t} = x_{t,o_t}] + \text{Cov}(\mathbb{E}[S_t|X_t]|X_{t,o_t} = x_{t,o_t}). \quad (31)$$

Преобразуем первое слагаемое, учитывая тот факт, что ковариация не зависит от реализации  $X_{t,o_t}$ .

$$\mathbb{E}[\text{Cov}(S_t|X_t)|X_{t,o_t} = x_{t,o_t}] = \text{Cov}(S_t|X_t) = \mathbf{\Gamma} - \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}. \quad (32)$$

Теперь преобразуем второе слагаемое:

$$\begin{aligned}\text{Cov}(\mathbb{E}[S_t|X_t]|X_{t,o_t} = x_{t,o_t}) &= \text{Cov}(\mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} X_t|X_{t,o_t} = x_{t,o_t}) \\ &= \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \text{Cov}(X_t|X_{t,o_t} = x_{t,o_t}) \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}.\end{aligned}$$

Таким образом:

$$\text{Cov}(S_t|X_{t,o_t} = x_{t,o_t}) = \mathbf{\Gamma} - \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma} + \mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} \text{Cov}(X_t|X_{t,o_t} = x_{t,o_t}) \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}. \quad (33)$$

Для полных наблюдений  $X_{t,o_t} = X_t$ , и, соответственно:

$$\text{Cov}(\mathbb{E}[S_t|X_t]|X_{t,o_t} = x_{t,o_t}) = \text{Cov}(\mathbb{E}[S_t|X_t]|X_t = x_t) = 0. \quad (34)$$

О выводе смешанных вторых моментов  $\mathbb{E}[X_t S_t^*|X_t = x_t]$  и  $\mathbb{E}[X_t S_t^*|X_{t,o_t} = x_{t,o_t}]$ :

$$\mathbb{E}[X_t S_t^*|X_t] = X_t \mathbb{E}[S_t^*|X_t] = X_t (\mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} X_t)^* = X_t X_t^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}, \quad (35)$$

$$\begin{aligned} \mathbb{E}[X_t S_t^*|X_{t,o_t} = x_{t,o_t}] &= \mathbb{E}[X_t \mathbb{E}[S_t^*|X_t]|X_{t,o_t} = x_{t,o_t}] = \mathbb{E}[X_t (\mathbf{\Gamma}^* \mathbf{A}^* \hat{\mathbf{R}}^{-1} X_t)^*|X_{t,o_t} = x_{t,o_t}] \\ &= \mathbb{E}[X_t X_t^* \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma}|X_{t,o_t} = x_{t,o_t}] = \mathbb{E}[X_t X_t^*|X_{t,o_t} = x_{t,o_t}] \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{\Gamma} \end{aligned}$$

В приведенных выше переходах используется закон полного условного математического ожидания и уже полученное выражение для  $\mathbb{E}[S_t^*|X_t]$ .

## Приложение №3. О реализации М-шага

Если антенная решетка является равномерной и линейной, удобно искать оптимальный  $u = \sin(\theta)$ , а затем находить  $\theta$  как  $\arcsin(u)$ . Для удобства введем новые обозначения:  $C = \mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*]$ ,  $D = \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]$ . Соответственно минимизации подлежит функция

$$\mathcal{Q}_1(u|u^{(\tau-1)}) = \|\mathbf{\Lambda}^{-1/2}(C - \tilde{A}(u)D)\|_F^2, \quad (36)$$

где

$$\tilde{A}(u) = \begin{bmatrix} 1 & 1 & \dots & 1 \\ e^{-2j\pi \frac{d}{\lambda} u_1} & e^{-2j\pi \frac{d}{\lambda} u_2} & \dots & e^{-2j\pi \frac{d}{\lambda} u_K} \\ \dots & \dots & \ddots & \vdots \\ e^{-2j\pi(L-1)\frac{d}{\lambda} u_1} & e^{-2j\pi(L-1)\frac{d}{\lambda} u_2} & \dots & e^{-2j\pi(L-1)\frac{d}{\lambda} u_K} \end{bmatrix}. \quad (37)$$

Для ускорения оптимизации можно оптимизировать не  $\mathcal{Q}_1(u|u^{(\tau-1)})$ , а суррогатную функцию  $\mathcal{G}(u|u^{(\tau-1)})$ , построенную так:

$$\mathcal{G}(u|u^{(\tau-1)}) = \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T (u - u^{(\tau-1)}) + \frac{1}{2} (u - u^{(\tau-1)})^T \mathbf{H} (u - u^{(\tau-1)}), \quad (38)$$

где

$$\mathbf{H} = \begin{bmatrix} \max(|\nabla_1 \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})|, \varepsilon) & 0 & \dots & 0 \\ 0 & \max(|\nabla_2 \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})|, \varepsilon) & \dots & 0 \\ \dots & \dots & \ddots & \vdots \\ 0 & 0 & \dots & \max(|\nabla_K \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})|, \varepsilon) \end{bmatrix}, \quad (39)$$

причем  $\nabla_1 \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) - i$ -я компонента градиента  $\mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})$ . Диагональная аппроксимация  $\mathbf{H}$  предотвращает слишком маленькие или слишком большие шаги, сохраняя численную стабильность. Теперь эту суррогатную функцию надо минимизировать (или, эквивалентно, найти направление шага  $\rho^{(\tau)} = u - u^{(\tau-1)}$ ):

$$\min_u \mathcal{G}(u|u^{(\tau-1)}). \quad (40)$$

Подставим  $u = u^{(\tau-1)} + \rho$ :

$$\mathcal{G}(u^{(\tau-1)} + \rho|u^{(\tau-1)}) = \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \rho + \frac{1}{2} \rho^T \mathbf{H} \rho. \quad (41)$$

Для нахождения минимума берем градиент по  $\rho$ :

$$\nabla_\rho \mathcal{G}(u^{(\tau-1)} + \rho|u^{(\tau-1)}) = \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \mathbf{H} \rho, \quad (42)$$

при  $\nabla_\rho \mathcal{G} = 0$  будет выполняться:

$$0 = \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \mathbf{H} \rho, \quad (43)$$

решаем относительно  $\rho^{(\tau)}$ :

$$\rho^{(\tau)} = -\mathbf{H}^{-1} \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}). \quad (44)$$

Важно заметить, что производная по направлению  $\nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \rho^{(\tau)}$  принимает вид:

$$\nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \rho^{(\tau)} = -\nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \mathbf{H}^{-1} \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}), \quad (45)$$

и поскольку  $\mathbf{H}$  — положительно определенная матрица, производная по направлению может принимать лишь неположительные значения, а значит направление  $\rho^{(\tau)}$  соответствует невозрастанию функции. Далее используется backtracking line search для гарантии невозрастания  $\mathcal{Q}_1$ :

$$u^{(\tau)} = u^{(\tau-1)} + \alpha_m \rho^{(\tau)}, \quad (46)$$

где  $\alpha_m$  – первое  $\alpha = \alpha_0 \beta^m, m \in \mathbb{N}$ , такое что  $\mathcal{Q}_1(u^{(\tau)}|u^{(\tau-1)}) \leq \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})$  для фиксированных  $\alpha_0 > 0, \beta \in (0; 1)$ .

В рамках программного комплекса реализуется мултистарт для поиска оптимального  $u$  на М-шаге. Указанный выше оптимизационный алгоритм запускается не только из точки  $u^{(\tau-1)}$ , но и из случайно выбранных допустимых точек, в том числе находящихся в окрестности  $u^{(\tau-1)}$ . Для выбранной начальной точки с индексом  $j$ , обновление оценки синусов угловых координат  $\tilde{u}^{(j)}$  может быть принято, только если:

1.  $\mathcal{Q}_1(\tilde{u}^{(j)}|u^{(\tau-1)}) > \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})$ ,
2.  $\mathcal{Q}_1(\tilde{u}^{(j)}|u^{(\tau-1)}) > \mathcal{Q}_1(\min(\tilde{u}^{(i)} : 1 \leq i \leq j-1)|u^{(\tau-1)})$ , при  $j > 1$ .

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

$$\mathbf{\Gamma}^{(\tau)} = \arg \min_{\mathbf{\Gamma}} \mathcal{Q}_2(\mathbf{\Gamma}|\mathbf{\Gamma}^{(\tau-1)}) = \mathbf{T} \left[ \log |\mathbf{\Gamma}| + \text{Tr}(\mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) \right]. \quad (47)$$

Определим точку, где производная данной функции принимает значение 0, и, таким образом, находим минимум функции:

$$\begin{aligned} \frac{\partial}{\partial \mathbf{\Gamma}} \log(\text{Det}(\mathbf{\Gamma})) &= \mathbf{\Gamma}^{-1}, \\ \frac{\partial}{\partial \mathbf{\Gamma}} \text{Tr}(\mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) &= -\mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \mathbf{\Gamma}^{-1}, \\ \frac{\partial \mathcal{Q}_2(\mathbf{\Gamma})}{\partial \mathbf{\Gamma}} &= \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \mathbf{\Gamma}^{-1}. \end{aligned}$$

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

$$\mathbf{O} = \mathbf{\Gamma}^{-1} - \mathbf{\Gamma}^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \mathbf{\Gamma}^{-1} \Rightarrow \mathbf{\Gamma}^{(\tau)} = \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*].$$

Предполагая, что сигналы некоррелированы, будем использовать лишь диагональное приближение матрицы, приравняв элементы вне главной диагонали к нулю:

$$\mathbf{\Gamma}^{(\tau)} = \mathcal{D} \left[ \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \right]. \quad (48)$$

Ввиду того, что М-шаг не максимизирует, а лишь улучшает УМО полного правдоподобия, представленный алгоритм является не точным ЕМ, а обобщенным (Generalized) ЕМ.