

ЕМ-алгоритм, стохастическая модель сигнала

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 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:

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

Сигналы, испускаемые источниками, шумы на сенсорах и наблюдения предполагаются стохастическими: $S_t \sim \mathcal{CN}(\mathbf{O}_{K \times 1}, \mathbf{\Gamma})$, $N_t \sim \mathcal{CN}(\mathbf{O}_{L \times 1}, \mathbf{\Lambda})$, $X_t \sim \mathcal{CN}(\mathbf{O}_{L \times 1}, A\mathbf{\Gamma}A^* + \mathbf{\Lambda})$, $t = 1, \dots, T$. 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.

Инициализация параметров

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

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

$$\nu \sim \mathcal{U}(-\pi; \pi); \quad (3)$$

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, $a \bmod b = a - b \cdot \lfloor \frac{a}{b} \rfloor$.

Начальную оценку ковариационной матрицы наблюдений $\hat{R}^{(0)}$ получаем на основе доступных данных. Начальную ковариацию сигналов $\Gamma^{(0)}$ получаем на основе метода наименьших квадратов с помощью $\theta^{(0)}, \hat{R}^{(0)}$.

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 $X_{t,o_t} = x_{t,o_t}, \Upsilon^{(\tau-1)}$:

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

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, 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.

Случай 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 $\hat{\mathbf{R}}$ на блоки, индуцированное этим разбиением множества индексов, оно имеет следующий вид:

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

где каждый блок определяется как

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

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 τ можно найти следующим образом:

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

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 $\mathbb{E}[X_t X_t^*|\mathcal{I}_t]$:

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

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, индуктировано разбиением множества индексов $\{1, \dots, L\}$ на множества o_t, m_t .

Параметры апостериорного распределения $P(S_t|\mathcal{I}_t, X_{t,m_t}), t = 1, \dots, T$ можно найти исходя из следующих формул, если $m_t \neq \emptyset$:

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

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

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

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

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

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

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

Случай 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 $P(S_t|\mathcal{I}_t), t = 1, \dots, T$:

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

При этом:

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

$$\mathbb{E}_{S_t|\mathcal{I}_t}[X_t S_t^*] = \mathbb{E}[X_t|\mathcal{I}_t] \cdot \mathbb{E}[S_t|\mathcal{I}_t]^* = x_t \cdot \mathbb{E}[S_t|\mathcal{I}_t]^*, \quad (13)$$

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

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:

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

Оценим угловые координаты источников θ :

$$\begin{aligned} \theta^{(\tau)} &= \arg \min_{\theta} Q_1(\theta|\theta^{(\tau-1)}) = \arg \min_{\theta} \left[-2 \text{Tr}(\Lambda^{-1} A(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*]) \right. \\ &\quad \left. + \text{Tr}(\Lambda^{-1} A(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] A^*(\theta)) \right] = \\ &\arg \min_{\theta} \|\Lambda^{-1/2} (\mathbb{E}_{(X_m, S)|\mathcal{I}}[XS^*] - A(\theta) \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*])\|_F^2. \end{aligned}$$

Эту задачу можно решить численно, подробности приведены в приложении 2.

Оценим ковариацию сигналов Γ :

$$\Gamma^{(\tau)} = \arg \min_{\Gamma} Q_2(\Gamma|\Gamma^{(\tau-1)}) = T \left[\log |\Gamma| + \text{Tr}(\Gamma^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) \right].$$

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

$$\begin{aligned} \frac{\partial}{\partial \Gamma} \log(\text{Det}(\Gamma)) &= \Gamma^{-1}, \\ \frac{\partial}{\partial \Gamma} \text{Tr}(\Gamma^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*]) &= -\Gamma^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \Gamma^{-1}, \\ \frac{\partial Q_2(\Gamma)}{\partial \Gamma} &= \Gamma^{-1} - \Gamma^{-1} \mathbb{E}_{(X_m, S)|\mathcal{I}}[SS^*] \Gamma^{-1}. \end{aligned}$$

Приравняем производную к нулю (функция по Γ выпукла):

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

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

$$\boldsymbol{\Gamma}^{(\tau)} = \mathcal{D} \left[\mathbb{E}_{(X_m, S) | \mathcal{I}}[SS^*] \right]. \quad (18)$$

Обновляем оценку ковариации наблюдений с учетом полученных оценок параметров:

$$\hat{\mathbf{R}}^{(\tau)} = \mathbf{A}(\theta^{(\tau)}) \boldsymbol{\Gamma} \mathbf{A}^*(\theta^{(\tau)}) + \boldsymbol{\Lambda}. \quad (19)$$

Результаты численного эксперимента

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. Об апостериорном распределении
сигналов, полученном на Е-шаге**

Приложение №2. О реализации М-шага

На практике, если A соответствует антенной решетки типа ULA, удобно искать оптимальный $u = \sin(\theta)$, а затем находить θ как $\arcsin(u)$. Для удобства введем новые обозначения: $C = \mathbb{E}_{(X_m, S) \mid \mathcal{I}}[XS^*]$, $D = \mathbb{E}_{(X_m, S) \mid \mathcal{I}}[SS^*]$. Соответственно минимизации подлежит функция

$$\mathcal{Q}_1(u|u^{(\tau-1)}) = \|\Lambda^{-1/2}(C - \tilde{A}(u)D)\|_F^2, \quad (20)$$

где

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

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

$$\mathcal{G}(u|u^{(\tau-1)}) = \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T(u - u^{(\tau-1)}) + \frac{1}{2}(u - u^{(\tau-1)})^T \mathbf{H}(u - u^{(\tau-1)}), \quad (21)$$

где

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

причем $\nabla_1 \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) - i$ -я компонента градиента $\mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})$. Теперь эту суррогатную функцию надо минимизировать (или, эквивалентно, найти направление шага $\rho^{(\tau)} = u - u^{(\tau-1)}$):

$$\min_u \mathcal{G}(u|u^{(\tau-1)}). \quad (22)$$

Подставим $u = u^{(\tau-1)} + \rho$:

$$\mathcal{G}(u^{(\tau-1)} + \rho|u^{(\tau-1)}) = \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \rho + \frac{1}{2} \rho^T \mathbf{H} \rho. \quad (23)$$

Для нахождения минимума берем градиент по ρ :

$$\nabla_\rho \mathcal{G}(u^{(\tau-1)} + \rho|u^{(\tau-1)}) = \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \mathbf{H} \rho, \quad (24)$$

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

$$0 = \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}) + \mathbf{H} \rho, \quad (25)$$

решаем относительно $\rho^{(\tau)}$:

$$\rho^{(\tau)} = -\mathbf{H}^{-1} \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}). \quad (26)$$

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

$$\nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \rho^{(\tau)} = -\nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)})^T \mathbf{H}^{-1} \nabla \mathcal{Q}_1(u^{(\tau-1)}|u^{(\tau-1)}), \quad (27)$$

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 :

$$u^{(\tau)} = u^{(\tau-1)} + \alpha_m \rho^{(\tau)}, \quad (28)$$

где α_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 $\alpha_0 > 0$, $\beta \in (0; 1)$.