Секция «Математика и механика»

Оценка сигналов со случайными компонентами в модели с последовательными экспериментами

Ордин Андрей Леонидович
Аспирант
Московский государственный университет имени М.В. Ломоносова, Механико-математический факультет, Москва, Россия
E-mail: andrey.ordin@yahoo.com

Задачи оценки разреженных сигналов часто возникают при анализе микромассивов генетических данных, а также в астрономических исследованиях. В классической постановке, как правило, рассматривается единственное наблюдение зашумленного вектора чистого сигнала. Нас же будет интересовать обобщение, приведенное в недавней работе [1]. А именно, пусть наблюдения имеют вид

$$y_{ij} = \sqrt{\phi_{ij}}\mu_i + \varepsilon_{ij}, \quad j = 1, \dots, k, \quad i = 1, \dots, n,$$

где k — число проводимых экспериментов, n — размерность вектора наблюдений. Предполагается, что каждому j-му эксперименту ($j = 1, \dots, k$) соответствуют множество $I_j \subseteq \{1, \dots, n\}$ и определенный уровень энергии $\mathcal{E}_j$, причем $\mathcal{E}_1 + \ldots + \mathcal{E}_k = n$. Энергия $\mathcal{E}_j$ равномерно распределяется по координатам вектора наблюдений таким образом, что $\phi_{ij} = \mathcal{E}_j/|I_j|$, если $i \in I_j \neq \varnothing$, и $\phi_{ij} = 0$ иначе. Сигнал μ считается разреженным в том смысле, что $\mathsf{P}(\mu_i = 0) = 1 - p_n$ для $i = 1, \dots, n$ и $p_n \in (0, 1)$, причем p_n мало. В статье [1] рассматривается задача определения множества $S_n(\mu)$ ненулевых компонент вектора μ, для чего авторы используют алгоритм под названием *Distilled Sensing*. Этот алгоритм определяет множества I_j следующим образом: $I_1 = \{1, \dots, n\}$, $I_{j+1} = I_j \cap \{i : y_{ij} > 0\}$, где $j = 1, \dots, k$. В настоящей работе мы показываем, как на основе такого алгоритма построить оценку сигнала μ, а также исследуем функцию потерь

$$R(\hat{\mu}, r) = \sup_{\mu \in \Theta_n(r)} \mathsf{E}\|\hat{\mu} - \mu\|^2,$$

где $\Theta_n(r) = \{\mu : \mathsf{P}(\mu_i = 0) = 1 - p_n, \mathsf{P}(\mu_i = \eta_i) = p_n, \eta_i \geq r\sqrt{\log n}$ для $i = 1, \dots, n\}$. Приведем наш основной результат.

Теорема. *Пусть $p_n \to 0$ при $n \to \infty$. Найдется такая константа r_0, что при достаточно большом числе экспериментов k для всякого $r > r_0$ выполнено соотношение*

$$\lim_{n \to \infty} \frac{R(\hat{\mu}, r)}{n} = 0.$$

Отметим, что в классических моделях аналогичные функции потерь имеют порядок $O(n)$. Таким образом, условия теоремы являются достаточными для того, чтобы в модели с последовательными экспериментами величина функции потерь была порядка $\bar{o}(n)$, $n \to \infty$.

Литература

1. J. Haupt, R. M. Castro, R. Nowak. Distilled sensing: Adaptive sampling for sparse detection and estimation // Information Theory, IEEE Transactions on 57.9 (2011): 6222-6235.