

Использование стохастической идентификации параметров при разделении смесей коррелированных глубоких гауссовских моделей

В.Е. Дементьев¹, А.Г. Ташлинский¹

¹Ульяновский государственный технический университет, Северный Венец 32, Ульяновск, Россия, 432027

Аннотация. Важной задачей при обработке разнородных данных является выявление их внутренних закономерностей, позволяющих определять скрытые объекты и области, а также оценить параметры этих областей. Для решения этой задачи предлагается использовать концепцию глубоких гауссовских моделей, которая даёт возможность описывать различные нестационарные процессы, близкие по своим свойствам к реальным многомерным сигналам. Объединение последовательности глубоких гауссовских процессов с марковскими случайными полями позволило описывать многомерные сигналы, используя корреляционные связи элемента последовательности с его окрестностью. Формулируются необходимые условия и определяются параметры для описания смеси коррелированных глубоких гауссовских моделей. Для оценивания параметров таких конструкций рассмотрена эффективность использования стохастической идентификации, в частности EM (SEM) алгоритма. Разработан необходимый для реализации метода математический аппарат и показана состоятельность предложенного подхода.

1. Введение

В последние годы значительно возросло число работ, посвященных анализу многомерных сигналов. Но при этом остается и ряд задач, отсутствие решения которых сдерживает дальнейшее развитие машинных методов обработки подобных сигналов. К числу таких задач относится и задача аналитического описания многомерных коррелированных массивов данных, сложность решений которой обусловлена пространственной неоднородностью и, как правило, многомерностью реальных данных. Эти свойства либо не учитываются большинством известных математических моделей, ориентированных на описание пространственно однородных сигналов [1, 2], либо их учёт при синтезе на основе моделей алгоритмов обработки и прогноза приводит к неоправданно высокой вычислительной сложности [2]. В настоящей работе для описания и обработки многомерных неоднородных случайных сигналов и их последовательностей предлагается использовать упорядоченное во времени множество глубоких гауссовских процессов. Особенностью этих процессов является вложенная структура, когда параметры модели сами являются реализациями процессов более низкого уровня. Это дает возможность вероятностного описания изменений поведения моделируемого многомерного сигнала, как в пространстве, так и во времени. В работах [3-5] показано, что использование глубоких гауссовских моделей позволяет достичь высокой адекватности при

описании реальных процессов. Так, в [4] исследована эффективность описания такими моделями нестационарных случайных процессов.

Важным частным случаем глубоких гауссовых моделей являются дважды стохастические модели [6-9], имеющие один скрытый слой, управляющий наблюдаемым слоем. В работах [8, 9] исследована адекватность моделей этого класса и показана возможность построения на их основе эффективных алгоритмов обработки. При этом к настоящему времени нерешенными являются две проблемы. Во-первых, работы в этой области посвящены в основном описанию одномерных процессов, хотя большинство реальных информационных массивов многомерны. Во-вторых, не решена проблема описания локальных корреляционных свойств и динамики их изменения в пределах одного слоя модели. Значение отсчета глубокого гауссова процесса определяется только скрытыми вложенными случайными величинами в этой точке и вспомогательными случайными добавками и не зависит от соседних отсчетов. Это существенно сокращает область применения подобных моделей. Для решения указанных проблем предлагается объединить последовательности глубоких гауссовых процессов с марковскими случайными полями, которые позволяют описывать многомерные сигналы, используя корреляционные связи элемента последовательности с его окрестностью.

Рассмотрим задачу в следующей постановке. Пусть задана последовательность из n векторных элементов (отсчетов) \bar{x}_i , $i = \overline{1, n}$, многомерного случайного процесса с размером вектора $N^{(1)}$, которую нужно описать дважды стохастической моделью. Будем считать, что моделируемая последовательность при этом разбивается на P сегментов с координатами отсчетов, соответствующих их границам: $1, \dots, n_1, \dots, n_P$, отличающихся в каждом l -ом, $l = \overline{1, P}$, сегменте значениями параметров дважды стохастической модели:

$$\bar{x}_{l,i} = \bar{\mu}_l^{(1)} + (\bar{\mu}_l^{(2)} + \bar{y}_{l,i-1})^T A_l \bar{x}_{l,i-1} + V_l^{(1)} \bar{\xi}_i^{(1)}, \quad (1)$$

где: $\bar{\mu}_l^{(k)}$, $k = 1, 2$ - вектора размером $N^{(k)}$, компонентами которых являются средние значения, принимаемые основной и вспомогательной векторными последовательностями на k -ом уровне модели в l -ом сегменте; $\bar{y}_{l,i} = R_l \bar{y}_{l,i-1} + V_l^{(2)} \bar{\xi}_i^{(2)}$ - вектор размером $N^{(2)}$, определяющий динамику изменения корреляционных свойств наблюдаемой последовательности; A_l - трехмерный тензор размером $N^{(1)} \times N^{(1)} \times N^{(2)}$, служащий для согласования размерностей наблюдаемой и скрытой последовательностей векторов; $V_l^{(k)}$ - матрица размером $N^{(k)} \times N^{(k)}$, определяющая ковариационные связи между элементами основной и вспомогательных векторных последовательностей на k -ом уровне модели в l -ом сегменте; R_l - матрица размером $N^{(2)} \times N^{(2)}$, задающая корреляционные свойства скрытой последовательности $\{\bar{y}_{l,i}\}$; $\bar{\xi}_i^{(1)}$, $\bar{\xi}_i^{(2)}$ - вектора, составленные из независимых гауссовских случайных величин с нулевым математическим ожиданием и единичной дисперсией.

Для модели (1), представляющей собой множества упорядоченных гауссовых процессов требуется найти совокупность параметров $\{n_l, R_l, \bar{\mu}_l^{(k)}, V_l^{(k)}\}$, $l = \overline{1, P}$, $k = 1, 2$, которую обозначим через θ . Ключевым отличием этой модели от рассмотренных, например, в работах [3,5] является наличие прямых авторегрессионных связей между соседними элементами в пределах одного уровня. Далее по аналогии терминологией работы [3] будем называть (1) моделью смеси коррелированных гауссовских глубоких процессов.

На рисунке 1 приведены простые примеры характерных частных случаев использования модели (1): на рисунке 1а - реализация 100 отсчетов одномерной ($N^1 = 1$) модели случайного процесса при $\mu_1^1 = 0$, $K = 1$, $P = 2$, $n_1 = 50$, используемой, в частности, при решении задач поиска разладки [10]; на рисунке 1б - реализация такого же объема одномерной дважды стохастической модели [8] с изменяющимся коэффициентом корреляции при $K = 2$, $P = 1$, $\mu_1^1 = 0$, $\mu_1^2 = 0.975$, $R_1 = 0.918$, $V_1^{(1)} = 1$, $V_1^{(2)} = 0.01$.

Визуальный анализ рисунков подтверждает нестационарный характер наблюдаемых процессов с выраженной границей между двумя его сегментами с разными корреляционными

функциями (рисунок 1а) и медленным изменением корреляционных характеристик (рисунок 1б).

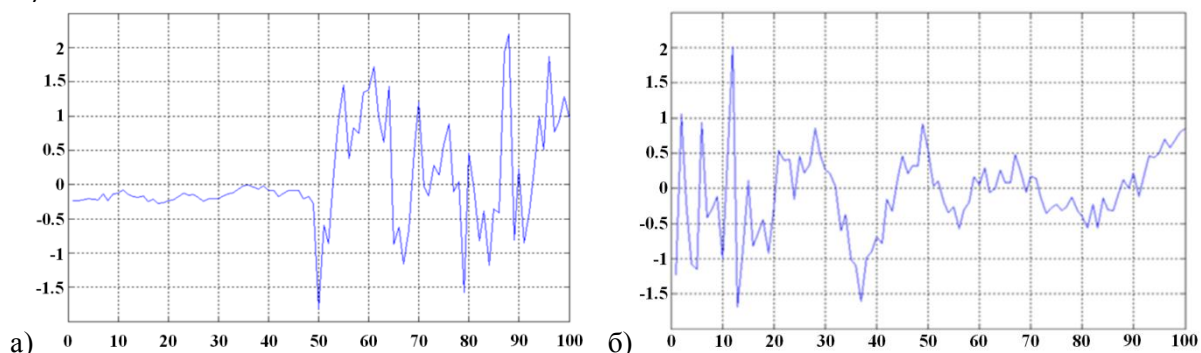


Рисунок 1. Пример простых реализаций смеси коррелированных глубоких гауссовских процессов.

2. Стохастическая идентификация параметров смеси коррелированных глубоких гауссовских процессов

Рассмотрим вопросы нахождения наилучших оценок $\hat{\theta}$ по имеющимся наблюдениям моделируемого процесса. При этом учтем, что условная плотность распределения вероятностей (ПРВ) $\varpi(x|y^{(1)}, \theta)$ является гауссовой и справедливы следующие равенства, непосредственно вытекающие из исходных систем:

$$\bar{y}_{l,i} = R_l \left[(\bar{x}_{l,i} - V_l^{(1)} \bar{\xi}_i^{(1)} - \bar{\mu}_l^{(1)}) (A_l \bar{x}_{l,i-1})^{-1} - \bar{\mu}_l^{(2)} \right] + V_l^{(2)} \bar{\xi}_i^{(2)}. \quad (2)$$

Предположим, что множества $\{\bar{x}_{l,i}, l = \overline{1, P}, i = \overline{1, n}\}$ и θ являются известными. Тогда $\{\bar{y}_{l,i}\}$ – представляют собой результат линейного преобразования относительно нормальных векторов $\bar{\xi}_i^{(1)}$ и $\bar{\xi}_i^{(2)}$. Соответственно условные ПРВ $\omega(Y = \{\bar{y}_{l,i}\} | X = \{\bar{x}_{l,i}\}, \theta)$ также являются нормальными. При этом вектор математических ожиданий соответствует

$$\bar{M}_i = M\{\bar{y}_{l,i} | X, \theta\} = R_p \left[(\bar{x}_{l,i} - \bar{\mu}_1^{(1)}) (A_l \bar{x}_{l,i-1})^{-1} - \bar{\mu}_1^{(2)} \right],$$

а четырехмерный ковариационный тензор:

$$V(i_1, i_2) = M\{(\bar{y}_{l_1, i_1} - \bar{M}_{i_1})(\bar{y}_{l_2, i_2} - \bar{M}_{i_2})^T | X, \theta\}$$

целесообразно представлять как двумерную диагональную матрицу, элементами которой при $i_1 \neq i_2$ являются нулевые матрицы размером $N^{(2)} \times N^{(2)}$, а при $i = i_1 = i_2$ – матрицы

$$T_l = R_l \left[V_l^{(1)} (A_l \bar{x}_{l,i-1})^{-1} + V_l^{(2)} \right]$$

размером $N^{(2)} \times N^{(2)}$. Тогда совместная условная ПРВ

$$\omega(Y | X, \theta) = \frac{1}{(2\pi)^{\frac{n}{2}} \prod_{i=1}^n \det[T_i]} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\bar{y}_{l,i} - \bar{M}_i) T_i^{-1} (\bar{y}_{l,i} - \bar{M}_i)^T\right). \quad (3)$$

Приведенные соотношения представляют собой по сути своеобразный «генератор», имитирующий неизвестный скрытый процесс при известных параметрах θ и отсчетах наблюдаемого процесса. Используя этот «генератор» можно имитировать нужное число реализаций $\{\tilde{Y} = \{\tilde{y}_{l,i}, l = \overline{1, P}, i = \overline{1, n}\}\}$ для обучения процедуры нахождения параметров θ . Для этих целей можно использовать аппарат стохастической идентификации, в частности ЕМ алгоритм [1], реализующий итеративную процедуру, максимизирующую функционал

$$\mathcal{L}(q, \theta) = \int q(Y) \log \frac{\omega(X, Y | \theta)}{q(Y)} dY,$$

где X – совокупность наблюдений, Y – множество скрытых параметров, q – некоторая неизвестная функция. На первом шаге, исходя из некоторой априорной информации или случайным образом, выбирается начальный вектор $\theta^{(0)}$. Далее на Е-шаге выполняется максимизация $\mathcal{L}(q, \theta)$ по q при заданном $\theta^{(0)}$. В работе [1] показано, что эта задача эквивалентна минимизации дивергенции Кульбака-Лейблера

$$q^{(0)}(Y) = \underset{q}{\operatorname{argmax}} \mathcal{L}(q, \theta^{(0)}) = \omega(Y|X, \theta^{(0)}).$$

На M-шаге выполняется поиск $\theta^{(1)}$, обеспечивающего максимум $\mathcal{L}(q^{(0)}(Y), \theta)$, т.е.

$$\theta^{(1)} = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(q^{(0)}(Y), \theta) = \underset{\theta}{\operatorname{argmax}} \int q^{(0)}(Y) \log \frac{\omega(X, Y|\theta)}{q^{(0)}(Y)} dY. \quad (4)$$

Далее оценка $\theta^{(1)}$ используется на E шаге.

Анализ показывает, что наиболее трудоемким является поиск $\theta^{(1)}$, обеспечивающего условие (4). Для снижения вычислительных затрат вместо поиска максимума функционала (4) можно применить существенно более простую процедуру определения $\theta^{(1)}$, обеспечивающего в среднем большие значения $\mathcal{L}(q^{(0)}(Y), \theta)$, чем даёт $\theta^{(0)}$. Для этого на основании полученной на E шаге ПРВ $\omega(Y|X, \theta^{(0)})$ выполняется генерация выборки \tilde{Y} и определяется значение стохастического градиента $\nabla_{\theta} \mathcal{L}(q^{(0)}(Y), \theta) = \frac{\partial}{\partial \theta} \log \frac{\omega(X, Y|\theta)}{q^{(0)}(\tilde{Y})}$, на основе которого находится очередное приближение:

$$\theta^{(1)} = \theta^{(0)} + \beta \nabla_{\theta} \mathcal{L}(q^{(0)}(Y), \theta),$$

где β задаёт приращения оцениваемых параметров θ .

3. Частные случаи стохастической идентификации параметров модели

Рассмотрим подробнее методику идентификации параметров модели для двух частных случаев. Вначале положим, что в модели (1) $P = 1$. Тогда X и Y – одномерные случайные последовательности, определяемые соотношениями

$$\begin{aligned} x_i &= \mu_1^{(1)} + (\mu_1^{(2)} + y_{1,i-1}^{(1)} x_{i-1}) + \sigma_1^2 \xi_i^{(1)}, \\ y_{1,i}^{(1)} &= r y_{1,i-1}^{(1)} + \sigma_2^2 \xi_i^{(2)}, i = \overline{1, n}, \end{aligned} \quad (5)$$

соответствуют простейшей дважды стохастической модели [9]. Найдем неизвестные параметры $\bar{\theta} = (\mu_1^{(1)}, \mu_1^{(2)}, r, \sigma_1^2, \sigma_2^2)^T$ модели (5).

Оценки параметров $\mu_1^{(1)}$, $\mu_1^{(2)}$ и σ_1^2 легко определяются из выражений, характеризующих статистические свойства последовательности (5):

$$\hat{\mu}_1^{(1)} = \frac{\sum_{i=1}^n x_i}{n}; \hat{\sigma}_1^2 = \frac{\sum_{i=1}^n (x_i - \hat{\mu}_1^{(1)})^2}{n-1}; \hat{\mu}_1^{(2)} = \frac{\sum_{i=1}^{n-1} (x_i - \hat{\mu}_1^{(1)})(x_{i+1} - \hat{\mu}_1^{(1)})}{\hat{\sigma}_1^2(n-1)}. \quad (6)$$

Для нахождения параметров r и σ_2^2 используем совместную условную ПРВ

$$\omega(X, Y|\theta) = \omega(X|Y, \theta) \omega(Y|\theta) = \frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp \left(- \sum_{i=1}^n 0.5 (\sigma_1^{-2} (x_i - K_i) x_{i-1})^2 + \sigma_2^{-2} (y_i - r y_{1,i-1}^{(1)})^2 \right),$$

где $K_i = \mu_1^{(1)} + (\mu_1^{(2)} - \tilde{y}_{1,i-1}^{(1)}) x_{i-1}$.

ПРВ $\omega(Y|X, \theta)$ для генерации обучающих выборок $\{\tilde{Y} = \{\tilde{y}_i, i = \overline{1, n}\}\}$ можно найти, представив уравнение для $y_{1,i}^{(1)}$ в (5) в виде:

$$y_{1,i}^{(1)} = r M_i - \frac{r \sigma_1^2}{x_{i-1}} \xi_i^{(1)} + \sigma_2^2 \xi_i^{(2)}, \quad (7)$$

откуда

$$\omega(Y|X, \theta) = \frac{1}{(2\pi)^{n/2} \prod_{i=1}^n \sqrt{D_i}} \exp \left(- \frac{1}{2} \sum_{i=1}^n \frac{h_i}{D_i} \right),$$

где $D_i = \frac{r \sigma_1^2}{x_{i-1}} + \sigma_2^2$, $M_i = r \frac{x_i - \mu_1^{(1)} - \mu_1^{(2)} x_{i-1}}{x_{i-1}}$, $h_i = (\tilde{y}_{1,i}^{(1)} - M_i)^2$ – обозначения, введенные для уменьшения громоздкости записи формул.

Тогда, обозначая через $A(X, \tilde{Y}, \theta)$ логарифм целевой функции качества идентификации:

$$A(X, \tilde{Y}, \theta) = \log \frac{\omega(X, \tilde{Y} | r, \sigma_2^2)}{q^{(0)}(\tilde{Y})} = \log \left(\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \right) - \frac{1}{2\sigma_1^2} \sum_{i=1}^n (x_i - K_i)^2 - \frac{1}{2\sigma_2^2} \sum_{i=1}^n z_i^2 + \frac{1}{2} \sum_{i=1}^n \frac{h_i}{D_i} + \frac{1}{2} \sum_{i=1}^n \log(D_i),$$

можно записать получаем псевдоградиенты для нахождения параметров r и σ_2^2 :

$$\frac{\partial A(X, \tilde{Y}, \theta)}{\partial r} = \frac{1}{2} \sum_{i=1}^n \frac{\sigma_1^2}{r \sigma_1^2 + x_{i-1} \sigma_2^2} - \frac{1}{\sigma_2^2} \sum_{i=1}^n (\tilde{y}_i - r \tilde{y}_{1,i-1}^{(1)}) \tilde{y}_{1,i-1}^{(1)} + \sum_{i=1}^n \frac{(\tilde{y}_{1,i}^{(1)} - D_i) \left(\frac{M_i}{r} \right)}{D_i} - \frac{1}{2} \sum_{i=1}^n \frac{h_i \sigma_1^2}{x_{i-1} D_i^2}.$$

$$\frac{\partial A(X, \tilde{Y}, \theta)}{\partial (\sigma_2^2)} = -\frac{n}{2\sigma_2^2} + \frac{1}{2\sigma_2^4} \sum_{i=1}^n (\tilde{y}_i - r\tilde{y}_{1,i-1}^{(1)})^2 - \frac{1}{2} \sum_{i=1}^n \frac{h_i}{D_i^2} + \frac{1}{2} \sum_{i=1}^n \frac{1}{D_i}$$

Поскольку в последнем выражении $\tilde{Y} = \{\tilde{y}_i\}_{i=1}^n$ - множество значений латентных случайных величин, полученных с помощью выражения (7), решение задачи определения оценки \hat{r} можно свести к следующей итеративной процедуре:

1. Задается начальное приближение оценок $\hat{\theta}^{(0)} = (\hat{r}^{(0)}, (\hat{\sigma}_2^2)^{(0)})^T$.
2. Для $\hat{\theta}^{(0)}$ имитируется множество выборок $\tilde{Y}_k, k = \overline{1, N}$.
3. Находятся псевдоградиенты $\nabla_r \mathcal{L}(q^{(0)}(Z), \theta^{(0)})$ и $\nabla_{\sigma_2^2} \mathcal{L}(q^{(0)}(Z), \theta^{(0)})$. При этом в качестве $\mathcal{L}(q^{(0)}(Z), \theta^{(0)})$ можно использовать медиану или среднее оценок $A(X, \tilde{Y}_k, \hat{\theta}^{(0)})$ логарифма целевой функции, $k = \overline{1, N}$.
4. Рассчитывается значение $\hat{\theta}^{(1)} = (\hat{r}^{(1)}, (\hat{\sigma}_2^2)^{(1)})^T = \hat{\theta}^{(0)} + \beta_1 \nabla_r \mathcal{L}(q^{(0)}(Z), r^0)$.
5. Полученные оценки всех параметров $\hat{\theta}^{(1)}$ используются в п. 1.

Коэффициент обучения Шаг β_1 псевдоградиентной релаксации можно оптимизировать, исходя из соображений, представленных в [11, 12].

На рисунке 2 представлен пример графиков сходимости оценок $\bar{\theta}^{(t)} = (\hat{r}^{(t)}, (\sigma_\rho^2)^{(t)})^T$ к истинным значениям $(0.9, 0.1)^T$.

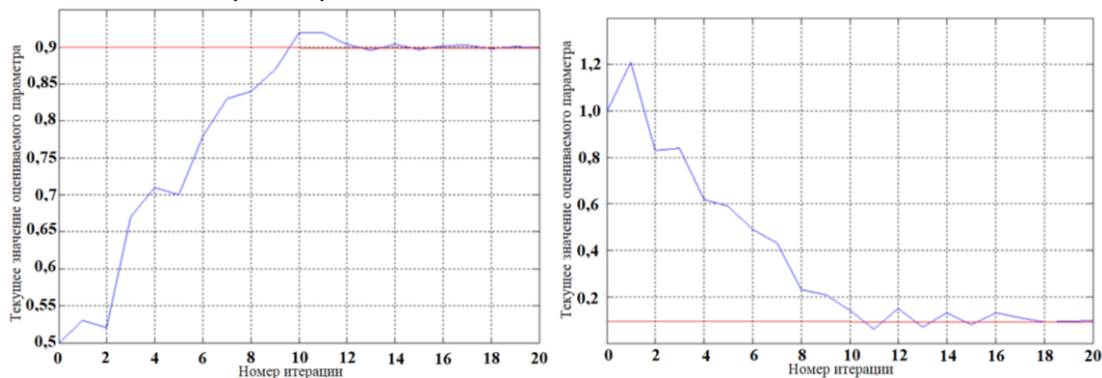


Рисунок 2. Процесс сходимости оценок $\bar{\theta}^t = (\hat{r}^{(t)}, \sigma_\rho^2)^T$.

Рассмотрим также случай, когда наблюдаемая последовательность описывается смесью из двух глубоких гауссовских моделей ($P = 2$). При этом

$$\bar{x}_{l,i} = \bar{\mu}_l^{(1)} + (\bar{\mu}_l^{(2)} + \bar{y}_{l,i-1})^T A_l \bar{x}_{l,i-1}^T + (\sigma_l^{(1)})^2 \bar{\xi}_i^{(1)}, \quad \bar{y}_{l,i} = R_l \bar{y}_{l,i-1} + (\sigma_l^{(1)})^2 \bar{\xi}_i^{(2)}, \quad l = \overline{1, 3}, \quad i = \overline{1, n}.$$

Тогда аналогично случаю, рассмотренному выше, можно найти

$$A(X, \tilde{Y}, \theta) = \log \frac{\omega(X, \tilde{Y} | \theta)}{q^{(0)}(\tilde{Y})} \text{ и } \frac{\partial A(X, \tilde{Y}, \theta)}{\partial r_1}, \frac{\partial A(X, \tilde{Y}, \theta)}{\partial r_2}, \frac{\partial A(X, \tilde{Y}, \theta)}{\partial (\sigma_1^{(2)})^2}, \frac{\partial A(X, \tilde{Y}, \theta)}{\partial (\sigma_2^{(2)})^2}.$$

Однако в силу дискретности n_2 найти аналитически производную $\frac{\partial A(X, \tilde{Y}, \theta)}{\partial n_2}$ не представляется возможным. Но можно использовать оценку этой производной через конечные разности, в частности:

$$\Delta A(X, \tilde{Y}, \theta) = A(X, \tilde{Y}, \bar{\theta}_2) - A(X, \tilde{Y}, \bar{\theta}_1).$$

Тогда определение $\hat{\theta}$ можно осуществить следующим блочно-координатным алгоритмом.

1. Задается начальное приближение оценок $\hat{\theta}^{(0)}$.
2. Для $\hat{\theta}^{(0)}$ имитируется множество выборок $\tilde{Y}_k, k = \overline{1, N}$.
3. Определяется $\hat{n}_2^{(1)} = \hat{n}_2^{(0)} + [\beta_2^{(1)} \hat{n}_2^{(0)} A(X, \tilde{Y}, \theta)]$, где $[\cdot]$ – операция округления.
4. С помощью соотношений, аналогичных (6), находятся оценки $\mu_1^{(1)}, \mu_1^{(2)}, (\sigma_1^2)^{(1)}, \mu_2^{(1)}, \mu_2^{(2)}, (\sigma_2^2)^{(1)}$.

5. Для оценки параметров r_1 , $(\sigma_1^{(2)})^2$, r_2 , и $(\sigma_2^{(2)})^2$ рассчитаются проекции $\nabla_{r_1} \mathcal{L}(q^{(0)}(Z), \hat{\theta}^{(0)})$, $\nabla_{(\sigma_1^{(2)})^2} \mathcal{L}(q^{(0)}(Z), \hat{\theta}^{(0)})$, $\nabla_{r_2} \mathcal{L}(q^{(0)}(Z), \hat{\theta}^{(0)})$ и $\nabla_{(\sigma_2^{(2)})^2} \mathcal{L}(q^{(0)}(Z), \hat{\theta}^{(0)})$ псевдоградиента целевой функции $\mathcal{L}(q, \theta)$.

6. Находятся оценки параметров r_1 , $(\sigma_1^{(2)})^2$, r_2 , и $(\sigma_2^{(2)})^2$.

7. Полученные оценки всех параметров $\hat{\theta}^{(1)}$ используются в п. 1

На рисунке 3 приведена реализация итеративного процесса поиска границы n_2 для процесса, изображенного на рисунке 1а.

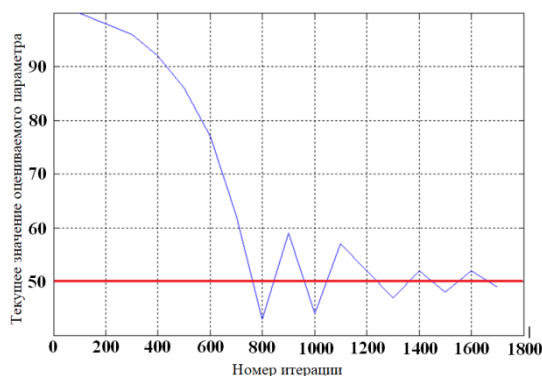


Рисунок 3. Процесс поиска границы между сегментами с разными параметрами глубокой гауссовской модели.

Анализ результатов, в частности и представленных на рисунках 2 и 3, позволяет утверждать о достаточно уверенной идентификации параметров смесей глубоких гауссовских процессов.

4. Выводы

В настоящей работе предложен математический аппарат, позволяющий описывать сигналы разной природы, представимые в виде комбинации нестационарных случайных процессов. На основе модифицированного стохастического EM алгоритма разработан метод идентификации параметров такой комбинации. В перспективе метод планируется обобщить для решения задачи сегментации многомерных сигналов, в частности спектрозональных изображений.

5. Благодарности

Результаты получены при поддержке грантов РФФИ №19-29-09048 и №18-47-730009.

6. Литература

- [1] Bouman, C.A. Model Based Imaging Processing / C.A. Bouman – West Lafayette: Purdue University, 2013. – 414 p.
- [2] Krasheninnikov, V. Multidimensional image models and processing / V. Krasheninnikov, K. Vasil'ev // Computer Vision in Control Systems-3. – 2018. – Vol. 135. – P. 11-64.
- [3] Viroli, C. Deep gaussian mixture models / C. Viroli, G. J. McLachlan [Electronic resource]. – Access mode: <https://arxiv.org/pdf/1711.06929.pdf> (01.12.2019).
- [4] Damianou, A. Deep Gaussian processes / A. Damianou, N. Lawrence // Artificial Intelligence and Statistics, 2013. – P. 207-215.
- [5] Salimbeni, H. Doubly stochastic variational inference for deep gaussian processes / H. Salimbeni, M. Deisenroth // Advances in Neural Information Processing Systems, 2017. – P. 4588-4599.
- [6] Woods, J.W. Image estimation using doubly stochastic Gaussian random field models / J.W. Woods, S. Dravida, R. Mediavilla // Pattern Analysis and Machine Intelligence. – 1987. – Vol. 9(2). – P. 245-253.
- [7] Шалыгин, А.С. Прикладные методы статистического моделирования / А.С. Шалыгин, Ю.А. Палагин – Л.: Машиностроение, 1986. – 320 с.

- [8] Dement'ev, V.E. Doubly stochastic models of images / V.E. Dement'ev, K.K. Vasil'ev, N.A. Andriyanov // *Pattern Recognition and Image Analysis*. – 2015. – Vol. 25(1). – P. 105-110. DOI: 10.1134/S1054661815010204.
- [9] Dementev, V.E. Filtration and restoration of satellite images using doubly stochastic random fields / V.E. Dementev, K.K. Vasiliev, N.A. Andriyanov // *CEUR Workshop Proceedings*. – 2017. – Vol. 1814. – P. 10-20.
- [10] Васильченко, С.Г. Алгоритм обнаружения моментов разладки случайной последовательности / С.Г. Васильченко // *Фундаментальная и прикладная математика*. – 2002. – Т. 8, № 3. – С. 655-665.
- [11] Tashlinskii, A.G. Pseudogradient Estimation of Digital Images Interframe Geometrical Deformations / A.G. Tashlinskii // *Vision Systems: Segmentation & Pattern Recognition*. – Vienna, Austria: I Tech Education and Publishing, 2007. – P. 465-494. DOI: 10.5772/4975.
- [12] Tashlinskii, A.G. The Specifics of Pseudogradient Estimation of Geometric Deformations in Image Sequences / A.G Tashlinskii // *Pattern Recognition and Image Analysis*. – 2008. – Vol. 18(4). – P. 701-706.

The use of stochastic parameter identification in the separation of mixtures of correlated deep Gaussian models

V.E. Dementiev¹, A.G. Tashlinskii¹

¹Ulyanovsk State Technical University, Severnii Venetz 32, Ulyanovsk, Russia, 432027

Abstract. An important task in the processing of different types of data is to identify different internal patterns in this data, allowing you to identify hidden objects and areas, as well as to evaluate the parameters of these areas. To solve this problem, it is proposed to use the concept of deep Gaussian models, which makes it possible to describe various non-stationary processes close in their properties to real multidimensional signals. Combining a sequence of deep Gaussian processes with Markov random fields made it possible to describe multidimensional signals using the correlation relationships of a sequence element with its vicinity. The paper formulates the necessary mathematical apparatus for describing mixtures of correlated deep Gaussian models. To identify the parameters of such constructions it is proposed to use a modification of the stochastic EM (EM) algorithm. The necessary relations for the implementation of this approach are determined and the consistency of the method is shown.