

Байесовский подход к анализу тренда и сезонности в марковской модели нестационарной регрессии

Даниил Кононенко

1. Институт Проблем Передачи Информации
127994, г. Москва, ГСП-4, Большой Каретный Переулок, 19, стр.1

2. DATADVANCE

105064, г. Москва, Садово-Черногрязская улица, 13/3

3. МФТИ

141700, г. Долгопрудный, Институтский переулок, 9
daniil.kononenko@datadvance.net

Аннотация

В работе рассматривается байесовский подход к задаче оценивания модели сигнала в рамках скрытой марковской модели. Предлагается способ моделирования тренда и сезонности в рамках линейной нормальной модели нестационарной регрессии. Задача решается с помощью обобщенной процедуры динамического программирования, работающей за линейное от длины сигнала время. Описывается вычислительный эксперимент на модельных данных.

1. Введение

Существует широкий класс задач анализа сигналов, в которых требуется для предъявленного сигнала $Y = (y_t, t = 1, \dots, N)$, $Y \in \mathcal{Y}$, подобрать наиболее подходящую модель \hat{X} из некоторого класса моделей $X \in \mathcal{X}$. Различают стационарные модели, передающие общую форму предъявленного сигнала в пределах всей области определения, и нестационарные модели, отражающие изменение некоторого локального свойства сигнала. Нестационарную модель сигнала следует искать либо в виде последовательности локальных моделей, либо в виде значений изменяющегося параметра некоторой общей локальной модели в каждой точке оси сигнала $X = (x_t, t = 1, \dots, N)$. Нестационарные модели рассматривают, например, при анализе сигнала речи [5], при анализе данных о доходностях ценных бумаг и оценивании изменяющегося состава портфеля инвестиционной компании [4–7], в задаче сглаживания сигнала [4].

2. Постановка задачи оценивания скрытой компоненты

2.1. Байесовский подход к оцениванию модели сигнала

Рассмотрим двухкомпонентный процесс $(X, Y) = [(x_t, y_t), t = 1, \dots, N]$, в котором наблюдаемая компонента $Y = (y_t, t = 1, \dots, N)$ есть подлежащий анализу сигнал, а скрытая компонента $X = (x_t, t = 1 \dots N)$, интерпретируется как его искомая нестационарная модель. Байесовский подход предполагает наличие вероятностного распределения $H(X, Y)$ на множестве всех реализаций двухкомпонентного процесса $(X, Y) \in \mathcal{X} \times \mathcal{Y}$. Также предполагается наличие априорного распределения $\Psi(X)$ на множестве моделей и условного распределения $\Phi(Y|X)$.

$$H(X, Y) = \Psi(X)\Phi(Y|X).$$

Определено также апостериорное распределение

$$P(X|Y) = \frac{H(X, Y)}{\int_{X' \in \mathcal{X}} H(X', Y) dX'} = \frac{\Psi(X)\Phi(Y|X)}{\int_{X' \in \mathcal{X}} \Psi(X')\Phi(Y|X') dX'}.$$

Пусть оператор оценивания скрытой компоненты по наблюдаемой имеет вид $\hat{X}(Y)$ и задана функция потерь $\lambda(X, \hat{X}) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$. Оптимальной будем считать оценку, минимизирующую средний риск ошибки — математическое ожидание функции потерь $r[\hat{X}] = M(\lambda(X, \hat{X}))$. Сингулярная функция потерь $\lambda(X, \hat{X}) = 1 - \delta(X, \hat{X})$ приводит к оптимизационной задаче максимизации логарифма апостериорной плотности распределения скрытого процесса [5]:

$$\hat{X}(Y) = \arg \max_{X \in \mathcal{X}^N} [\log \Psi(X) + \log \Phi(Y|X)]. \quad (1)$$

2.2. Скрытая марковская модель сигнала

В рамках этой модели скрытый процесс будем считать марковским, т.е. для условных плотностей распределения выполнено

$$\psi_t(x_t|x_1, \dots, x_{t-1}) = \psi_t(x_t|x_{t-1}).$$

Также будем считать, что имеется некоторая априорная информация о значении процесса в момент $t = 0$, выраженная в виде безусловной плотности распределения $\psi_0(x_0)$. Тогда полагая, что $X = (x_t, t = 0, 1, \dots, N)$, получаем

$$\Psi(X) = \psi_0(x_0) \prod_{t=1}^N \psi_t(x_t|x_{t-1}). \quad (2)$$

Будем предполагать, что каждое из мгновенных условных распределений наблюдаемого процесса определяется единственным значением скрытого процесса в текущий момент времени:

$$\varphi_t(y_t|y_1, \dots, y_{t-1}, X) = \varphi_t(y_t|y_1, \dots, y_{t-1}, x_t). \quad (3)$$

Обозначим через $Y^t = (y_s, s \leq t)$ фрагмент наблюдаемого процесса, заканчивающийся справа моментом t , рассматриваемым как последний момент наблюдения. Будем называть фильтрационной условную плотность распределения

$$p_{t|t}(x_t) = p_t(x_t|Y^t), \quad (4)$$

а интерполяционной — условную плотность распределения

$$p_{t|N}(x_t) = p_t(x_t|Y). \quad (5)$$

Задачей фильтрации назовем задачу нахождения оптимальной оценки скрытой переменной \hat{x}_t , если известен только фрагмент наблюдаемого процесса Y^t , а задачей интерполяции — задачу нахождения оптимальной оценки скрытой переменной \hat{x}_t , если известен весь наблюдаемый процесс Y . Из определения видно, что решение задачи интерполяции есть искомого оптимальное значение скрытых переменных.

2.3. Постановка задачи динамического программирования

Согласно предположению о связи наблюдаемого случайного процесса со скрытым процессом (3) имеем

$$\log \varphi(Y|X) = \sum_{t=1}^N \log \varphi_t(y_t|Y^{t-1}; x_t) = \sum_{t=1}^N \eta_t(x_t|Y^t).$$

В силу предположения о марковском свойстве скрытого процесса (2) получаем

$$\begin{aligned} \log \Psi(X) &= \log \psi_0(x_0) + \sum_{t=1}^N \log \psi_t(x_t|x_{t-1}) = \\ &= \log \psi_0(x_0) + \sum_{t=1}^N \log \eta_t(x_t|Y^t) + \sum_{t=1}^N \log \gamma_t(x_{t-1}, x_t). \end{aligned}$$

Приходим к следующей оптимизационной задаче для оценки скрытого процесса:

$$\begin{cases} X(Y) = [\hat{x}_0(Y), \dots, \hat{x}_N(Y)] = \arg \max_{X \in \mathcal{X}} J(x_0, \dots, x_N|Y), \\ J(x_0, \dots, x_N|Y) = \sum_{t=0}^N \eta_t(x_t|Y^t) + \sum_{t=1}^N \gamma_t(x_{t-1}, x_t). \end{cases} \quad (6)$$

Функции $\eta_t(x_t|Y^t)$ будем называть узловыми, а функции $\gamma_t(x_{t-1}, x_t)$ — функциями связи.

3. Оценивание скрытого марковского процесса

Задачи оптимизации вида (6) детально рассмотрены в работе [8]. Они решаются с помощью метода динамического программирования, изначально изложенного в [1].

Рассмотрим последовательность функций

$$J_t(x_0, \dots, x_t|Y^t) = \sum_{s=0}^t \eta_s(x_s|Y^s) + \sum_{s=1}^t \gamma_t(x_{s-1}, x_s). \quad (7)$$

Определим функции Беллмана одного аргумента

$$J_{t|t}(x_t) = \min_{x_0, \dots, x_{t-1}} J_t(x_0, \dots, x_t|Y^t).$$

Теорема 3.1. *Функции Беллмана связаны рекуррентным соотношением*

$$J_{t|t}(x_t) = \eta_t(x_t|Y^t) + \min_{x_{t-1} \in \mathcal{X}} [\gamma_t(x_{t-1}, x_t) + J_{t-1|t-1}(x_{t-1})]. \quad (8)$$

Эта теорема позволяет определять функции Беллмана последовательно для возрастающих моментов времени, начиная с известной нулевой функции. Непосредственно из определения (7) следует, что точка минимума функции Беллмана $J_{t|t}(x_t)$ дает решение задачи фильтрации. Решение задачи фильтрации в последней точке $\hat{x}_{N|N}$ является также решением задачи интерполяции.

Определим последовательность маргинальных функций

$$J_{t|N}(x_t) = \min_{x_s, s \neq t} J(x_0, \dots, x_N|Y). \quad (9)$$

Теорема 3.2. *Маргинальные функции удовлетворяют следующему обратному рекуррентному соотношению:*

$$\begin{aligned} J_{t-1|N}(x_{t-1}) &= \min_{x_t \in \mathcal{X}} \left\{ J_{t|N}(x_t) + [J_{t-1|t-1}(x_{t-1}) + \gamma_t(x_{t-1}, x_t)] \right\} - \\ &= \min_{x'_{t-1} \in \mathcal{X}} [J_{t-1|t-1}(x'_{t-1}) + \gamma_t(x'_{t-1}, x_t)]. \end{aligned}$$

По своему определению (9) маргинальные функции дают решение задачи интерполяции

$$\hat{x}_{t|N} = \arg \min_{x_t \in \mathcal{X}} J_{t|N}(x_t).$$

Таким образом, последовательно определяя функции Беллмана и маргинальные функции, можно решить задачу за линейное от длины сигнала время, что оказывается важно для многих практических приложений. Доказательства теорем 3.1 и 3.2 можно найти, например, в [5].

3.1. Линейная нормальная модель

В качестве априорной информации о скрытой компоненте естественно взять предположение о «близости» смежных локальных моделей x_{t-1} и x_t . Одной из самых простых и в то же время часто встречаемых на практике форм такой близости является линейная связь. Пусть в каждый момент времени скрытой компоненте отвечает вектор скрытых переменных $\mathbf{x}_t \in \mathbb{R}^n$, наблюдаемой компоненте — скалярная величина $y_t \in \mathbb{R}$.

$$\begin{cases} \mathbf{x}_t = \mathbf{V}_t \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, & t = 2, \dots, N; \\ y_t = \mathbf{c}_t^T \mathbf{x}_t + \xi_t, & t = 1, \dots, N; \\ \mathbf{V}_t \in \mathbb{R}^{n \times n}, M(\boldsymbol{\varepsilon}_t) = \mathbf{0}, M(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_s^T) = 0, \text{ при } t \neq s, \\ \text{Cov}(\boldsymbol{\varepsilon}_t) = M(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T) = \boldsymbol{\sigma}^2 \mathbf{U}_t^{-1}; \\ \mathbf{c}_t \in \mathbb{R}^n, M(\xi_t) = 0, D(\xi_t) = \sigma_y^2, \\ M(\xi_{t_1} \xi_{t_2}) = 0, \text{ при } t_1 \neq t_2, M(\boldsymbol{\varepsilon}_t \xi_s) = \mathbf{0}. \end{cases} \quad (10)$$

Здесь \mathbf{U}_t^{-1} — ковариационная матрица "внутреннего" шума.

Будем полагать, что аддитивные шумы $\boldsymbol{\varepsilon}_t$ и ξ_t нормальны

$$\psi(\boldsymbol{\varepsilon}_t) = \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2 \mathbf{U}_t^{-1}), \quad \varphi(\xi_t) = \mathcal{N}(0, \sigma_y^2),$$

а также нормально распределено случайное начальное значение скрытого процесса

$$\psi_0(\mathbf{x}_0) = \mathcal{N}(\hat{\mathbf{x}}_{0|0}, \mathbf{B}_{0|0}^{-1}) \quad (11)$$

Очевидно, что условные распределения $\psi_t(\mathbf{x}_t | \mathbf{x}_{t-1})$ и $\varphi_t(y_t | \mathbf{x}_t)$ также нормальны:

$$\begin{aligned} \psi_t(\mathbf{x}_t | \mathbf{x}_{t-1}) &= \mathcal{N}(\mathbf{V}_t \mathbf{x}_{t-1}, \boldsymbol{\sigma}^2 \mathbf{U}_t^{-1}), \\ \varphi_t(y_t | \mathbf{x}_t) &= \mathcal{N}(\mathbf{c}_t^T \mathbf{x}_t, \sigma_y^2). \end{aligned}$$

Тогда для узловых функций и функций связи имеем

$$\eta_t(x_t | Y^t) = \left\{ -\log \sigma - \frac{1}{2} \log(2\pi) \right\} - \frac{1}{2\sigma^2} (y_t - \mathbf{c}_t^T \mathbf{x}_t)^2, \quad (12)$$

$$\gamma_t(x_{t-1}, x_t) = \left\{ -\log \sigma - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{U}_t^{-1}| \right\} - \frac{1}{2\sigma^2} (\mathbf{x}_t - \mathbf{V}_t \mathbf{x}_{t-1})^T \mathbf{U}_t (\mathbf{x}_t - \mathbf{V}_t \mathbf{x}_{t-1}). \quad (13)$$

Процедура фильтрации [5] имеет вид

$$\begin{cases} \hat{\mathbf{x}}_{t|t} = \mathbf{V}_t \hat{\mathbf{x}}_{t-1|t-1} + \mathbf{B}_{t|t}^{-1} \mathbf{c}_t (y_t - \mathbf{c}_t^T \mathbf{V}_t \hat{\mathbf{x}}_{t-1|t-1}), \\ \mathbf{B}_{t|t} = \mathbf{c}_t \mathbf{c}_t^T + (\mathbf{V}_t \mathbf{B}_{t-1|t-1} \mathbf{V}_t^T + \mathbf{U}_t^{-1})^{-1}. \end{cases} \quad (14)$$

Процедура интерполяции [5]:

$$\begin{cases} \hat{\mathbf{x}}_{t|N} = \hat{\mathbf{x}}_{t|t} + \mathbf{H}_t (\hat{\mathbf{x}}_{t+1|N} - \mathbf{V}_{t+1} \hat{\mathbf{x}}_{t|t}), \\ \mathbf{B}_{t|N}^{-1} = \mathbf{H}_t \mathbf{B}_{t+1|N}^{-1} \mathbf{H}_t^T + (\mathbf{V}_{t+1}^T \mathbf{U}_{t+1} \mathbf{V}_{t+1} + \mathbf{B}_{t|t}^{-1})^{-1}, \\ \mathbf{H}_t = \mathbf{B}_{t|t}^{-1} \mathbf{V}_{t+1}^T (\mathbf{V}_{t+1} \mathbf{B}_{t|t}^{-1} \mathbf{V}_{t+1}^T + \mathbf{U}_{t+1}^{-1})^{-1}. \end{cases} \quad (15)$$

4. Модели тренда и сезонности

В данном разделе предлагается подход к построению скрытой модели сигнала с учетом тренда и сезонности. Необходимо использовать марковские модели тренда и сезонности, чтобы сохранить линейную вычислительную сложность всей процедуры.

Пусть трендовая часть наблюдаемого сигнала выражается полиномом степени не выше k : $g(t) = a_k t^k + \dots + a_1 t + a_0$, $k < N$.

Утверждение 4.1. $\forall t \ g_{t+k+1} = \alpha_0 g_t + \alpha_1 g_{t+1} + \dots + \alpha_k g_{t+k} = \sum_{i=0}^k \alpha_i g_{t+i}$, причем коэффициенты $\alpha_0, \dots, \alpha_k$ не зависят от коэффициентов конкретного полинома a_0, \dots, a_k , а только от его степени k и задаются выражением

$$\alpha_i = (-1)^{k-i} C_{k+1}^i. \quad (16)$$

Тогда можно записать

$$\begin{pmatrix} g_t \\ \vdots \\ g_{t+k} \end{pmatrix} = V^s \begin{pmatrix} g_{t-k-1} \\ \vdots \\ g_{t-1} \end{pmatrix}, \quad (17)$$

где матрица V^s размера $[(k+1) \times (k+1)]$ зависит только от степени полинома k .

Пусть сезонная (периодическая) часть наблюдаемого сигнала выражается l гармониками: $h(t) = a_0 + \sum_{i=1}^l a_i \cos(\Omega_i t) + b_i \sin(\Omega_i t)$, $\Omega_1, \dots, \Omega_l$ фиксированы. Рассмотрим случай $\Omega_i = 2\pi i$.

Утверждение 4.2. $\forall t \ h_{t+2l+1} = \alpha_0 h_t + \alpha_1 h_{t+1} + \dots + \alpha_{2l} h_{t+2l} = \sum_{i=0}^{2l} \alpha_i h_{t+i}$, причем коэффициенты $\alpha_0, \dots, \alpha_{2l}$ не зависят от коэффициентов a_0, a_1, \dots, a_l и b_1, \dots, b_l , а только от l .

Тогда можно записать

$$\begin{pmatrix} h_t \\ \vdots \\ h_{t+2l} \end{pmatrix} = V^h \begin{pmatrix} h_{t-2l-1} \\ \vdots \\ h_{t-1} \end{pmatrix}, \quad (18)$$

где матрица V^h размера $[(2l+1) \times (2l+1)]$ зависит только от l .

Объединим линейную нормальную модель (10) с моделями тренда (17) и сезонности (18). Для это

в каждый момент времени t добавим к скрытому вектору \mathbf{x}_t вектор тренда $\mathbf{g}_t = (g_t^1, \dots, g_t^{k+1})^T$ и вектор сезонности $\mathbf{h}_t = (h_t^1, \dots, h_t^{2l+1})^T$. Размерность полученного вектора скрытых переменных $\mathbf{z}_t = (\mathbf{x}_t^T, \mathbf{g}_t^T, \mathbf{h}_t^T)^T$ равна $\tilde{n} = n + k + 2l + 2$, здесь n — размерность \mathbf{x}_t .

Компоненты тренда и сезонности будем считать незашумленными, тогда для них справедливы соотношения (17) и (18). Так как мы рассматриваем тренд и периодическую составляющую наблюдаемого сигнала y_t , то соответствующие координаты вектора \mathbf{c}_t единичны:

$$c_t^{n+1} = c_t^{n+2} = \dots = c_t^{\tilde{n}} = 1.$$

Отсутствию шума в компонентах тренда и сезонности соответствует нулевой шумовой вектор, поэтому шумовой вектор для всей скрытой компоненты

$$\boldsymbol{\varepsilon}_t = \begin{pmatrix} \varepsilon_t^x \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^{\tilde{n}}.$$

Получаем следующую модель:

$$\begin{cases} \mathbf{z}_t = \mathbf{V}_t \mathbf{z}_{t-1} + \boldsymbol{\varepsilon}_t, & t = 2, \dots, N; \\ y_t = \mathbf{c}_t^T \mathbf{z}_t + \xi_t = (\mathbf{c}_t^T)^T \mathbf{x}_t + g_t^1 + \dots + h_t^{2l+1} + \xi_t. \end{cases} \quad (19)$$

Здесь

$$\mathbf{V}_t = \begin{pmatrix} \mathbf{V}_t^x & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^g & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{V}^h \end{pmatrix} \in \mathbb{R}^{\tilde{n} \times \tilde{n}},$$

$$\mathbf{x}_t = \mathbf{V}_t^x \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t^x, \quad \mathbf{V}_t^x \in \mathbb{R}^{n \times n},$$

$$\mathbf{g}_t = \mathbf{V}^g \mathbf{g}_{t-1}, \quad \mathbf{V}^g \in \mathbb{R}^{(k+1) \times (k+1)}, \quad (20)$$

$$\mathbf{h}_t = \mathbf{V}^h \mathbf{h}_{t-1}, \quad \mathbf{V}^h \in \mathbb{R}^{(2l+1) \times (2l+1)}. \quad (21)$$

Замечание 4.1. В данной модели одному моменту времени t соответствуют $(k+1)$ значение тренда и $(2l+1)$ значение сезонности. Однако такая модель имеет физический смысл в часто встречающемся на практике случае, когда наблюдаемый сигнал $Y = (y_t, t = 1, \dots, N)$ есть дискретизация некоторой непрерывной величины $y(t)$. Тогда, например, наличие $(k+1)$ значения тренда можно интерпретировать как разбиение интервала $(t-1, t)$ на $(k+1)$ часть. Таким образом, трендовая и сезонная компоненты определены на более точной дискретизации по времени, то есть на более мелком разбиении всего интервала времени $(1, N)$.

5. Вычислительный эксперимент

Был проведен вычислительный эксперимент на модельных данных. Рассматривалась трехмерная скрытая компонента $\mathbf{x}_t = (x_t^1, x_t^2, x_t^3)$, квадратичный тренд и одна гармоника сезонности. В соответствии

с моделью порождения данных (19) были сгенерированы скрытые сигналы \mathbf{x}_t , \mathbf{g}_t , \mathbf{h}_t и наблюдаемый сигнал y_t . Метод оценки скрытой компоненты реализован в среде MATLAB.

Результаты эксперимента представлены на Рис. 1-4. Данный эксперимент показывает что один и тот же наблюдаемый сигнал может быть образован в результате разных комбинаций трендовой, сезонной и зашумленной компоненты. В связи с этим важно точно задавать априорные предположения о тренде и сезонности для конкретной практической задачи в виде априорного математического ожидания и ковариационной матрицы скрытой компоненты (11). Также целесообразно вводить регуляризационный множитель в функционал (6).

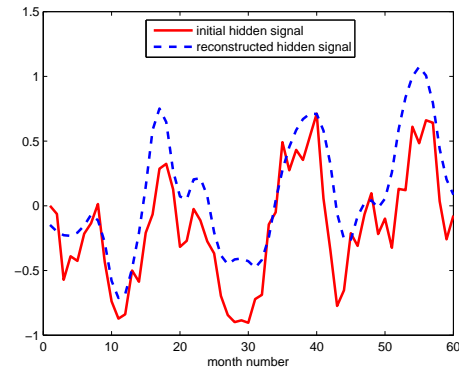


Рис. 1. Зашумленная скрытая компонента

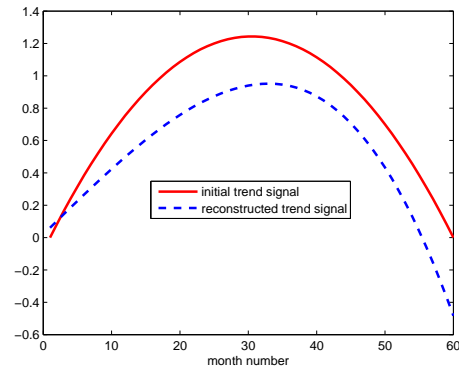


Рис. 2. Скрытая компонента тренда

6. Заключение

Предложен способ моделирования тренда и сезонности в рамках линейной нормальной модели нестационарной регрессии. Предложенный метод позволяет строить скрытую модель наблюдаемого сигнала с учетом тренда и сезонности за линейное от длины сигнала время.

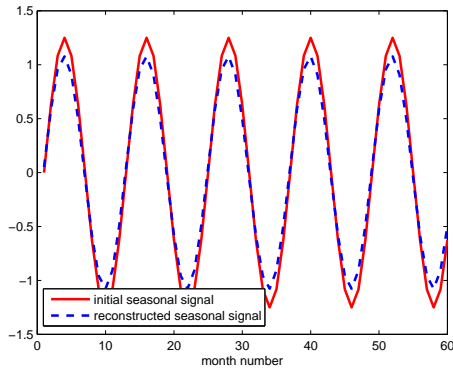


Рис. 3. Скрытая компонента сезонности

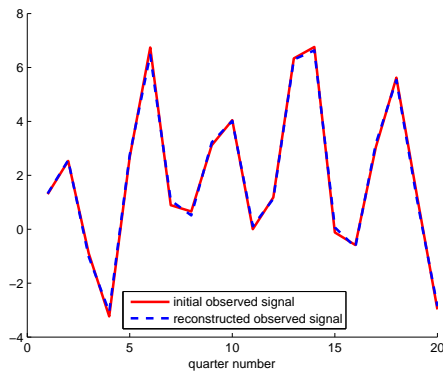


Рис. 4. Наблюдаемый сигнал

Список литературы

- [1] Richard Bellman, *Динамическое программирование*, М.: Мир, 1960.
- [2] Р. П. Федоренко, *Введение в вычислительную физику*, МФТИ, 1994.
- [3] R. Kalaba and L. Tesfatsion, *Time-varying linear regression via flexible least squares*, Computers Math. Applic., 1989, pp. 1215–1245.
- [4] О. В. Красоткина, *Алгоритмы оценивания моделей нестационарных сигналов при наличии ограничений*, Ph.D Thesis, Тульский государственный университет, 2003.
- [5] В. В. Моттль, *Статистические методы анализа сигналов и массивов упорядоченных данных*, unpublished manuscript.
- [6] Vadim Mottl, Olga Krasotkina, Michael Markov and Ilya Muchnik, *Dynamic analysis of hedge funds*, proceedings of the 8th IASTED International Conference on Financial Engineering and Applications. MIT, Cambridge, Massachusetts, USA, October 9-11, 2006.
- [7] Vadim Mottl, Olga Krasotkina, Michael Markov and Ilya Muchnik, *Time-varying regression model with unknown time-volatility for nonstationary signal analysis*, proceedings of the 8th IASTED International Conference on Signal and Image Processing, Honolulu, Hawaii, USA, August 14-16, 2006.

- [8] Ilya Muchnik and Vadim Mottl, *Bellman Functions on Trees for Segmentation, Generalized Smoothing, Matching and Multi-Alignment in Massive Data Sets*, DIMACS, Technical Report 98-15, February, 1998.