

*Шаг 8.* Занести категорию в профиль пользователя, включив соответствующую тройку  $(C_i, W_i, \alpha_i)$  в профиль согласно схеме (см. рисунок).

*Шаг 9.* Если уровень изменчивости  $\alpha_i > \alpha_0$ , где  $\alpha_0$  – заданная величина, то увеличить текущий вес категории  $C_i$  на величину  $\Delta W_i$ :

$$W_i = W_i + \Delta W_i.$$

*Шаг 10.* Отсортировать последовательность троек  $(C_i, W_i, \alpha_i)$  в профиле по порядку убывания веса  $W_i$ .

*Шаг 11.* Сохранить получившийся профиль.

Таким образом, в данной статье был рассмотрен алгоритм непрерывной корректировки модели (профиля) пользователя. Для успешного построения алгоритма предложена схема организации профиля пользователя в виде множества троек вида (категория интересов  $C_i$ , текущий вес категории  $w_i$ , уровень изменчивости  $\alpha_i$ ). При этом профиль делится на две группы (два подпрофиля): краткосрочный и долгосрочный – для учета краткосрочных и долгосрочных интересов пользователя (в том числе неявных). Кроме того, было введено понятие временного измерения в латент-

ном семантическом пространстве, что позволило адаптировать методологию PLSA для непрерывной оценки изменений интересов пользователя.

Применение предложенного алгоритма для подстройки модели в процессе ее работы с использованием неявной обратной связи, приближает нас к созданию высококачественных и эффективных поисковых систем с персонализированным интерфейсом.

#### Библиографические ссылки

1. Hoffman T. Unsupervised Learning by Probabilistic Latent Semantic Analysis // Machine Learning. 2008. Vol. 42. P. 177–196.
2. Salton G., McGill M. J. Introduction to Modern Information Retrieval. New York : McGraw-Hill, 1993.
3. Indexing by Latent Semantic Analysis / S. Deerwes, S. Dumasis, G. Furnas et al. // J. of the Amer. Soc. for Inform. Science. 1990. Vol. 41. P. 391–407.
4. Hoffman T. Probabilistic Latent Semantic Indexing // Proc. of the 22nd Annu. Intern. ACM SIGIR Conf. on Research and Development in Inform. Retrieval. Berkeley, Calif., 2009. P. 50–57.

М. В. Карасьева

#### APPLICATION OF PLSA METHODOLOGY FOR ADAPTIVE CORRECTION OF USER MODEL

*The paper considers the algorithm of the continuous model (profile) correction. The initial data are the initial profile and the previous inquiry history. The PLSA (Probabilistic Latent Semantic Analysis) methodology is used in the algorithm. To achieve the object in view, the term temporary latent semantic space is introduced.*

*Keywords: probabilistic latent semantic analysis, user model, user profile.*

© Карасева М. В., 2012

УДК 519.254

В. С. Кедрин, О. В. Кузьмин

#### ВЫДЕЛЕНИЕ ОСЦИЛЛИРУЮЩИХ И ТРЕНДОВЫХ КОМПОНЕНТ НА БАЗЕ КРИТЕРИАЛЬНОЙ МОДИФИКАЦИИ СИНГУЛЯРНОГО АНАЛИЗА

*Рассматривается методика выделения осциллирующих колебательных и трендовых составляющих при анализе сложных нестационарных процессов, протекающих в реальных сложных системах.*

*Ключевые слова: нестационарная система, временной ряд, сингулярное разложение, графические критерии качества, осциллирующие компоненты, трендовые компоненты.*

Одним из интенсивно развивающихся теоретических подходов к моделированию сложных процессов является использование непараметрических моделей динамических систем в виде набора элементарных характеристик (временных выборок), позволяющих по экспериментальным данным входа-выхода выявить динамические свойства и оценить состояние исследуемой системы. В этой связи особо актуальными становятся исследования, посвященные анализу факторов, влияющих на состоятельность практических рекомендаций в области идентификации и синтеза цифровых непараметрических моделей систем, в ко-

торых протекают процессы с ярко выраженными нестационарными свойствами. Примерами таких систем являются современные электроэнергетические системы и системы товарных и фондовых рынков.

В настоящее время существует несколько подходов к решению проблем, связанных с анализом нестационарной динамики.

Один из этих подходов основан на методах, заимствованных из нелинейной динамики [1–5]. Однако данный подход является ограниченным, так как он подразумевает постоянство оператора эволюции системы, в силу чего нестационарность, вызванная

изменением структурного и количественного состава элементов системы и их связей, не может быть адекватно исследована и определена.

Альтернативным подходом к анализу нестационарной динамики является разделение временных выборок, характеризующих данную динамику, на квазистационарные участки с последующей их классификацией. В этом случае целью анализа является отслеживание состояния выделенных участков. Само состояние квазистационарных участков может быть определено:

- с помощью оценки спектрального состава на базе оконного преобразования Фурье [6; 7]. Однако применительно к анализу нестационарного поведения системы такая оценка также является ограниченной в силу чувствительности преобразования Фурье к локальным скачкам и пикам функции. Поэтому применение этого преобразования является неэффективным для процессов, которые могут быть представлены функциями, являющимися суммами периодических компонент с эволюционирующими во времени частотой и амплитудой. Кроме того, преобразование Фурье является взаимно однозначным преобразованием, в связи с чем вся случайная составляющая, которая содержится в гладкой функции, переходит в спектр. Для уменьшения случайности выполняют переход к спектральной плотности и используют спектральные окна. При этом возникают сложности с интерпретацией спектров в случае детерминированной функции, так как трудно заранее определить долю случайности в исследуемом процессе;

- с помощью специальных статистических тестов, основанных на анализе характера распределений выделяемых участков [1; 8; 9]. Однако такой анализ может привести к неконтролируемой ошибке [10]. Еще одним существенным недостатком статистической классификации является неопределенность взаимодействия отдельных осциллирующих составляющих модели, связанная с трудностями учета относительных фазовых сдвигов комплексных выборок компонент.

Очевидно, что в силу указанных ограничений использование любого из этих подходов будет недостаточно эффективным для обоснованного анализа нестационарной системы, характеризующейся интенсивными и значительными изменениями показателей, а также непостоянством составляющих движения.

Таким образом, задачу непараметрического моделирования сложных динамических систем нельзя считать решенной и необходим поиск новых эффективных методов, позволяющих сформировать оценку динамического состояния системы в определенном режиме ее работы.

**Аппарат сингулярного разложения.** В последнее время активно развиваются новые спектральные методы анализа, с помощью которых можно получать более полную информацию о нестационарном временном ряде. Так, вейвлет-преобразование благодаря подвижным частотно-временным окнам одинаково хорошо выявляет как низкочастотные, так и высокочастотные характеристики. При этом можно выделить

как определенную пространственную (временную) частоту, так и ее локализацию в физическом пространстве (времени).

Перспективным направлением в области исследования процессов, обладающих сложной структурой, также является метод, основанный на сингулярном разложении временного ряда на простейшие аддитивные составляющие, который позволяет исследовать структуры временных рядов [11; 12].

Суть этого метода заключается в преобразовании одномерной выборки нестационарного процесса в матрицу развертки с помощью однопараметрической процедуры элементов ряда и сингулярного разложения этой матрицы на основании фундаментального соотношения

$$\mathbf{A} = \mathbf{U} \times \mathbf{W} \times \mathbf{V}^T, \quad (1)$$

где  $\mathbf{U}$  – унитарная матрица размером  $M \times M$ ;  $\mathbf{V}$  – унитарная матрица размером  $N \times N$ ;  $\mathbf{W}$  – матрица размером  $M \times N$ , на главной диагонали которой находятся сингулярные неотрицательные числа, расположенные в порядке убывания, а все недиагональные элементы равны нулю.

Из полученного набора главных сингулярных чисел, лежащих на диагонали матрицы  $\mathbf{W}$  в соотношении (1), по условию значимости (убыванию модуля) выбирается такое их количество, по которому может быть восстановлена совокупность аддитивных составляющих, сумма которых с допустимой ошибкой совпадает с выборкой исходного процесса. Более подробно аппарат сингулярного разложения описан в [11–15].

Анализ, выполненный в [13], показал следующие преимущества сингулярного разложения перед другими известными дискретными методами спектрального анализа:

- набор функций разложения порождается самой исследуемой функцией процесса  $f(t)$  и длиной окна  $M$ ;

- длина строки (окна) сингулярной матрицы развертки  $M$  позволяет легко варьировать качество и состав выделяемых составляющих;

- существует возможность управляемого восстановления исходного процесса по интерпретируемым компонентам, в отличие от практически однозначных компонент фурье- и вейвлет-преобразований;

- для реальных временных рядов отсутствует граничный эффект по параметру сдвига, определяемого, например, жесткой фиксацией набора вейвлет-функций;

- представление отдельной собственной сингулярной функции в виде линейного фильтра показывает, что он обладает не комплексной, как в случае преобразования Фурье, а действительной частотной характеристикой, что снимает проблемы, связанные с моделированием фазовых сдвигов между составляющими.

Таким образом, несмотря на достаточную простоту реализации, метод сингулярного разложения матрицы развертки (СРМР) имеет ряд важных свойств, позволяющих применять его для исследования сложных динамических процессов.

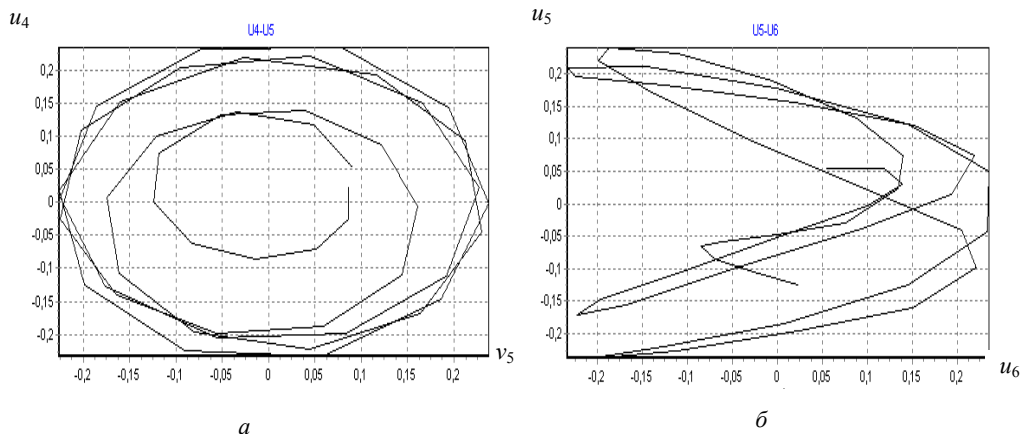


Рис. 1. Графики, образованные парами собственных векторов: *а* – высокое качество соответствия идеальному гармоническому сигналу; *б* – низкое качество соответствия идеальному гармоническому сигналу

**Формирование графических критериев качества выделения составляющих.** Для оценки большинства сложных систем необходим отбор колебательных составляющих из полученной в результате сингулярного анализа системы компонент разложения, так как они характеризуют физическую природу происходящих в системе волновых процессов, откуда возникает задача отбора осциллирующих составляющих. Для решения данной задачи на основе анализа фигур, образуемых в пространстве собственных векторов сингулярной матрицы развертки (рис. 1), была выполнена формализация графического критерия для определения соответствия выделяемых сингулярных компонент идеальному гармоническому сигналу.

Следует отметить, что степень близости фигур к кругу (рис. 1, *а*), определяемая по эллипсоидному движению точки в пространстве собственных векторов и по отношению сторон ограничивающего прямоугольника, характеризует уровень соответствия компонент идеальному гармоническому сигналу. При этом для оценки близости выделяемых компонент к идеальному гармоническому сигналу можно использовать критерий уровня отклонения колебаний *C*:

$$C = \max(\Delta A_i) : i = 1, l-1, \quad (1)$$

где  $\Delta A_i$  – приращение между двумя соседними колебаниями (рис. 2);  $l$  – количество колебаний.

Помимо определения осциллирующих компонент важное значение для оценки развития сложной системы имеет выделение компонент, характеризующих плавные тренды, которые косвенно могут оценивать:

- наличие в системе процессов, характерный временной масштаб которых больше длительности интервала наблюдения;
- влияние внешних факторов воздействия;
- дрейф параметров системы.

Для выделения плавных трендов при анализе сингулярного разложения предложен критерий уровня гладкости *H*, который характеризует угол приращения  $\beta$ , т. е. изменение значения производной в *i*-й точке анализируемой выборки (рис. 3):

$$H = \max(\beta_i) : i = 1, n-1. \quad (2)$$

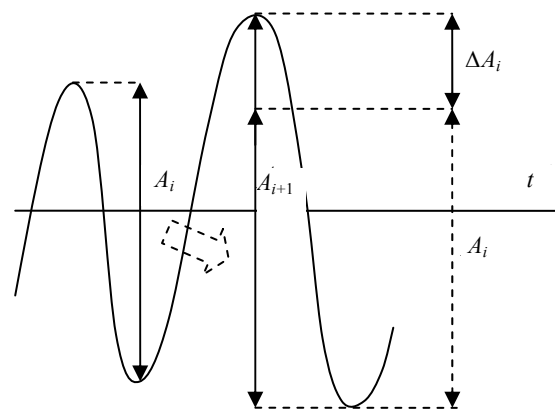


Рис. 2. Определение базовых параметров для критерия уровня отклонения колебаний

Этот критерий позволяет выявлять плавную трендовую динамику в исследуемом процессе исходя из заданного значения  $\beta$ .

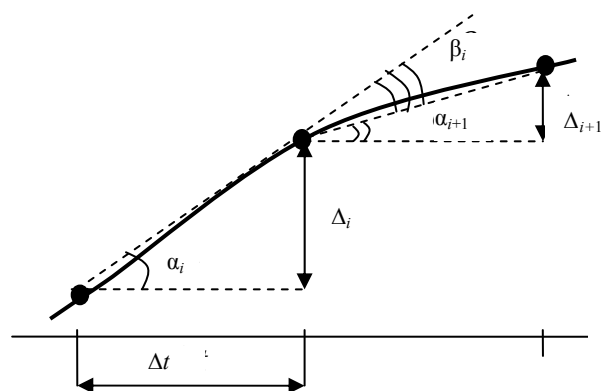


Рис. 3. Определение базовых параметров для критерия уровня гладкости

**Методика критериальной модификации сингулярного разложения.** Представленные на рис. 1–3 графические критерии позволяют автоматизировать отбор выделяемых сингулярных составляющих, близ-

ких к гармоническим и нелинейным плавным сигналам. Это обстоятельство было положено в основу модифицированной методики сингулярного разложения, с помощью которой можно оптимальным образом разделять исходный процесс до заданного уровня случайных составляющих и избегать получения искаженных составляющих движения. Алгоритм модифицированной методики сингулярного анализа с использованием графических критериев качества представлен ниже (рис. 4).

Применение стандартного метода сингулярного анализа предполагает, что параметр  $M$  должен быть зафиксирован [14; 15]. В то же время для выделения осциллирующих составляющих такое требование является неоптимальным, поскольку выделение конкретных гармонических составляющих зависит от параметра  $M$ , который должен быть кратен периоду колебаний [14; 15]. Поэтому для применения предложенных графических критериев качества целесообразно производить выделение гармонических компонент при изменении параметра  $M$ . В этом случае формируется множество трендовых и осциллирующих компонент и возникает задача построения различных моделей комбинаторных конфигураций для определения вероятности присутствия выделенных свойств у исходного процесса, которая в данной статье не рассматривается, но предпосылки для ее решения можно найти в работе [16].

**Пример 1.** Рассмотрим процесс изменения напряжения  $U(t)$  на Жигулевской гидроэлектростанции (рис. 5), который характеризуется сложной нестационарной динамикой, что может объясняться многосвязностью и многомерностью сложной электроэнергетической системы. При этом диапазон рабочих режимов функционирования этой системы значительно расширяется.

Так, в результате применения графических критериев качества для выделения осциллирующих составляющих при задании конкретного уровня отклонения колебаний  $C$  и параметра метода сингулярного разложения  $M = 100$  для представленного на рис. 1 процесса были получены следующие результаты (табл. 1).

Результаты модифицированного сингулярного анализа с использованием графических критериев качества для рассматриваемого в примере 1 процесса (см. рис. 5) представлены в табл. 2.

По данным табл. 2 можно сделать вывод о том, что применение модифицированной критериальной методики сингулярного разложения позволяет повысить качество выделения гармонических компонент для сложных реальных электроэнергетических процессов и, следовательно, качество оценок динамического состояния современных электроэнергетических систем.

**Пример 2.** В табл. 3 представлены результаты модифицированного сингулярного анализа с использованием графических критериев качества для процессов, протекающих на фондовом рынке.

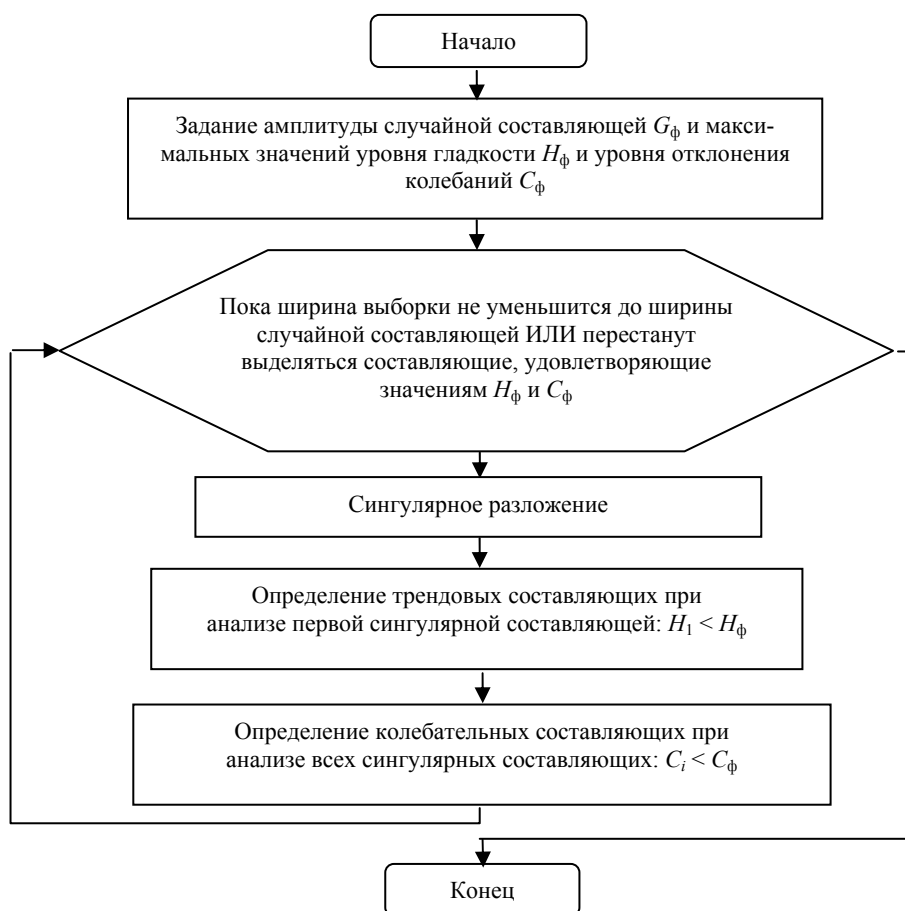


Рис. 4. Блок-схема методики модифицированного сингулярного анализа



Рис. 5. Процесс изменения напряжения на Жигулевской ГЭС (200 значений) по состоянию на 12 января 2007 г.

Таблица 1

**Результаты применения метода стандартного сингулярного анализа для выделения гармонических компонент**

Уровень отклонения колебаний $C$ , %	20	30	40	50	60
Число гармонических компонент	0	0	0	2	5
Уровень от сигнала, %	0	0	0	3,51	6,11

Таблица 2

**Результаты применения методики модифицированного сингулярного анализа для выделения гармонических компонент**

Уровень отклонения колебаний $C$ , %	20	30	40	50	60
Число гармонических компонент	0	2	6	16	82
Уровень от сигнала, %	0	3,26	8,30	7,09	10,99

Таблица 3

**Исследование структурного состава процесса изменения валютного курса Eur/USD**

Интервал времени	Трендовые составляющие			Гармонические составляющие			Случайная составляющая Уровень, %
	Количество	Уровень, %	Средний уровень критерия $H$	Количество	Уровень, %	Средний уровень критерия $K$	
С 5 июня 2005 г. по 29 мая 2007 г.	5	66,16	22,4	42	24,24	38,33	9,57
С 4 января 2007 г. по 30 мая 2007 г.	2	74,67	16,5	97	16,06	38,64	10,51
С 23 июня 2007 г. по 30 июня 2007 г.	1	14,04	29	33	74,55	38,11	11,41

Анализ данных табл. 3 показывает, что на долгосрочном и среднесрочном интервалах преобладает трендовая динамика, на краткосрочном интервале возрастает роль периодических и случайных факто-

ров, что отражает структурный состав выделяемых сингулярных компонент. Таким образом, более стабильный и предсказуемый характер имеют макроэкономические процессы на долгосрочном и среднесроч-

ном интервалах. Этот вывод согласуется с теоретическими положениями, сформулированными специалистами в области макроэкономических систем.

Таким образом, расширение и модификация классического алгоритма сингулярного разложения позволяют создать новую, достаточно перспективную методику анализа динамической нестационарности, с помощью которой можно получать оптимальный состав компонент, характеризующихся осциллирующими и трендовыми составляющими движения, имеющими предсказуемый характер и простое математическое описание. В частности, применение предложенной методики повышает качество выделения свободных колебательных составляющих в процессах, протекающих в сложных электроэнергетических системах, и помогает оценить состояние макроэкономических процессов на разных интервалах времени.

### Библиографические ссылки

1. Manuca R., Savit R. Stationary and Nonstationary Time Series Analysis // *Physica D*. 1996. Vol. 99. P. 134–161.
2. Kennel M. B. Statistical Test for Dynamical Nonstationarity in Observed Time Series Data // *Phys. Rev. E*. 1997. Vol. 56. P. 316.
3. Yu Dejin, Lu Weiping, Harrison R. G. SpaceTime-Index Plots for Probing Dynamical Nonstationarity // *Phys. Letters A*. 1998. Vol. 250. P. 323–327.
4. Measuring Nonstationarity by Analyzing the Loss of Recurrence in Dynamical Systems / C. Rieke, K. Stemickel, R. G. Andrzejak, et al. // *Phys. Rev. Letters*. 2002. Vol. 88, № 24. P. 24–27.
5. Schreiber T. Detecting and Analyzing Nonstationarity in a Time Series Using Nonlinear Cross Predictions // *Phys. Rev. Letters*. 1997. Vol. 78. P. 843.
6. Mallat S., Zhang Z. Matching Pursuit with Time-Frequency Dictionaries // *IEEE Trans. on Signal Processing*. 1993. Vol. 41. P. 3397–3415.
7. Горюнова Н. К., Дойников А. Н., Терешко Л. А. Влияние параметров режима электрической системы на ее частотные характеристики // *Тр. Ленингр. политехн. ин-та*. 1981. № 380. С. 26–30.
8. Гребенюк Е. А. Обнаружение изменений свойств нестационарных случайных процессов // *Автоматика и телемеханика*. 2003. № 12. С. 25–41.
9. Гребенюк Е. А. Анализ и оперативная диагностика систем, описываемых нестационарными случайными процессами // *Пробл. управл.* 2003. № 4. С. 23–29.
10. Жиров М. В., Макаров В. В., Солдатов В. В. Идентификация и адаптивное управление технологическими процессами с нестационарными параметрами. М. : Изд-во Моск. гос. техн. ун-та им. Н. Э. Баумана, 2011.
11. Дойников А. Н., Кедрин В. С., Сальникова М. К. Методика синтеза математических моделей рядов макроэкономических показателей на основе алгоритмов сингулярного разложения // *Вестн. Иркут. гос. техн. ун-та*. 2006. № 2. С. 138–142.
12. Дойников А. Н., Кедрин В. С., Сальникова М. К. Моделирование нестационарных процессов с использованием алгоритмов их сингулярного разложения // *Науч.-техн. ведомости С.-Петербур. гос. политехн. ун-та*. 2006. № 5. С. 143–147.
13. Кедрин В. С. Спектральные методы анализа в задачах оценки состояния сложных энергетических систем // *Вестн. Иркут. регион. отд-ния Акад. наук высш. шк. России*. 2008. №1 (13). С. 193–198.
14. Голяндина Н. Э. Метод «Гусеница»-SSA : анализ временных рядов : учеб. пособие. СПб. : Изд-во С.-Петербур. гос. ун-та, 2004.
15. Главные компоненты временных рядов: метод «Гусеница» / под ред. Д. Л. Данилова, А. А. Жиглявского. СПб. : Пресском, 1997.
16. Кузьмин О. В. Обобщенные пирамиды Паскаля и их приложения. Новосибирск : Наука. Сиб. издат. фирма РАН, 2000.

V. S. Kedrin, O. V. Kuzmin

### DISCRIMINATION OF OSCILLATE AND TREND COMPONENTS ON THE BASIS OF CRITERIA MODIFICATION OF SINGULAR ANALYSIS

*The authors consider a technique of discrimination of oscillating and trend components in the analysis of complex non-stationary processes in real complex systems.*

*Keywords: time-dependent system, time series, singular value decomposition, image quality criteria, the oscillating components, trend components.*

© Кедрин В. С., Кузьмин О. В., 2012